DOCUMENT RESUME

ED 462 757 EA 031 611

AUTHOR Mitchell, Douglas E.; Mitchell, Ross E.
TITLE The Impact of California's Class Size Reduction Initiative on Student Achievement: Detailed Findings from Eight School Districts.
PUB DATE 1999-00-00
NOTE 55p.
PUB TYPE Reports - Evaluative (142)
EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Academic Achievement; *Class Size; Classroom Environment; Elementary Secondary Education; *Small Classes
IDENTIFIERS *California; Reform Efforts

ABSTRACT

This report presents a comprehensive preliminary analysis of how California's Class Size Reduction (CSR) initiative has impacted student achievement during the first 2 years of implementation. The analysis is based on complete student, classroom, and teacher records from 26,126 students in 1,174 classrooms from 83 schools in 8 Southern California school districts. The evidence reviewed supports nine broad conclusions, including the following: (1) Evaluating the impact of CSR was made difficult because many other important initiatives were being simultaneously pursued; (2) rapid implementation of CSR placed substantial stresses on school facilities, created an intense demand for new teachers and encouraged a shift to year round school calendars; (3) impact from CSR is small but positive; (4) benefits of CSR experience were not evenly distributed among student groups; and (5) because class-size reduction was so deeply entangled with student, school, and teacher variables, it was impossible to disentangle the factors influencing student achievement with usual post-hoc exploratory data analysis. Recommendations include "staying the course" with class size until its full effect can be analyzed and documented; supporting work that establishes appropriate explanatory frameworks for interpreting the relationship between class size and student achievement; and continuing the search for school reform and improvement policies that can offset educational challenges created by poverty, non-English home language, and student ethnicity. (Contains 39 references.) (RT)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 462 757

The Impact of California's Class Size Reduction Initiative on Student Achievement: Detailed Findings from Eight School Districts

Douglas E. Mitchell
Ross E. Mitchell

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

R. Mitchell

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

EA031611

The Impact of California's Class Size Reduction Initiative on Student Achievement: Detailed Findings from Eight School Districts

Douglas E. Mitchell
Ross E. Mitchell
California Educational Research Cooperative
University of California, Riverside

Executive Summary

This report presents a comprehensive preliminary analysis of how California's Class Size Reduction (CSR) initiative has impacted student achievement during the first two years of implementation. The analysis is based on complete student, classroom and teacher records from 26,126 students in 1,174 classrooms from 83 schools in 8 Southern California school districts. The data include reading, mathematics and language test scores from the Stanford Achievement Test (Version 9 – SAT-9) collected through California's STAR testing program. Also analyzed are 34 variables covering student demographics, school assignments, classroom contexts, and teacher characteristics. The evidence reviewed supports nine broad conclusions and leads to five recommendations to education professionals and policy makers.

Conclusion #1: CSR is massive, expensive and adopted in conjunction with a complex array of other new policy initiatives aimed at improving California school performance. Evaluating the impact of this initiative is made particularly difficult by the fact that so many other important initiatives are being simultaneously pursued.

At a direct cost exceeding $2.3 billion in the first two years of implementation, CSR is the most expensive reform of public education ever undertaken in California (California Department of Education 1999). There are many reasons for believing that CSR may be helpful to public education. Improving student achievement is certainly its most important goal, however. Thus, student achievement effects of CSR implementation are the focus of this report. CSR was not adopted as an experiment or as a test of how much it could contribute to student performance, but was implemented comprehensively and on very short notice. Moreover, CSR was adopted at the same time as revisions in teacher preparation, mandates for reforming bilingual education, development of new curriculum frameworks and materials, adoption of a new statewide test, development of a new performance accountability system and numerous other policies whose effects cannot be precisely estimated. It may never be possible to know with certainty how much this initiative has contributed to student learning.

Conclusion #2: Rapid implementation of California's CSR initiative placed substantial stresses on school facilities, created an intense demand for new teachers, and encouraged a shift to Year Round school calendars to accommodate enrollment growth and reduced size classes.

EA031611

These stresses are quite likely to mean that CSR is functioning differently during its first few years of operation than can be expected in the years ahead. Schools hired many more teachers who are not fully credentialed and who lack training comparable to the average teacher in the years immediately prior to CSR implementation. An earlier study by the California Educational Research Cooperative documented a sharply elevated frequency of first-year and not fully qualified teachers serving in reduced size classes (Ogawa and Stine 1998).

Conclusion #3: School officials were faced with tough decisions regarding the sequence of CSR implementation and the allocation of opportunities to participate in reduced size classes on the part of teachers and students.

As a consequence of the choices made, students and teachers were definitely not randomly distributed among large and small classes. Of 34 variables examined in this study, only student gender did not significantly relate to whether students were assigned to large or small classes for one or both of the first two years of CSR implementation. Since academic achievement is influenced by multiple layers of demographic influences, classroom assignment variables, school and classroom contexts, and teacher characteristics, any effort to evaluate the impact of CSR must carefully attend to the imbalances in student and teacher participation.

Conclusion #4: Implementation biases responsible for differences in student and teacher participation in reduced size classes were strikingly different in the first and second years of CSR implementation.

Students in reduced size classes during the first year were more likely to be from ethnic minority groups, from poor neighborhoods and attending year round schools than those first participating in the second year of implementation. Students who did not have access to reduced size classes until the second year were more likely to be new to the district in 1998, and to come from English speaking homes.

Conclusion #5: Statistical analyses revealed that biases in CSR participation are sufficiently strong that knowing the demographic, school assignment and teacher characteristics of any given student makes it possible to substantially predict whether they were in small or large classes for one or both years.

Specifically, multiple discriminant analysis of CSR implementation biases improves by more than 35 percent our ability to predict their CSR experience. This means, quite simply, that achievement differences between the large and small classes created by California's CSR may be, to a substantial degree, determined by differences in who has participated, rather than how class size itself affects learning.

Conclusion #6: The factors associated with the biases in student participation in various CSR implementation alternatives are, themselves, much more strongly related to student achievement than is class size reduction.

Twenty-five of the 34 variables examined in this study were at least as powerful as CSR experience in predicting student achievement. Of these variables, student poverty, gender, ethnicity, home language, special education certification and transiency are two to twenty times as powerful as CSR experience in predicting student achievement. Additionally, teacher contract status, ethnicity, education level and gender are from two to ten times as powerful as CSR experience in predicting student achievement. As a result, relatively small biases in the assignment of students or teachers to small classes can create outcome differences that are as large or larger than the CSR effect.

> Conclusion #7: Nevertheless, after controlling for all of the available biasing factors, there remains a small positive impact from CSR on student achievement as measured by the Stanford-9 achievement test. The CSR impact varies from year to year, however, and is not consistent across the reading, mathematics and language subtests of the SAT-9.

After statistically removing the effects of the known biasing variables, CSR experience during the first two years of implementation accounted for about a 1 NCE point gain on the 1998 SAT-9 tests. This amount of achievement gain is approximately as much as would be expected to result from about two weeks of additional student maturation and instruction. The CSR contribution accounts for about one-tenth of one percent of all student achievement variation, whereas the other variables analyzed in this report account for 35.2 to 38.6 percent of student achievement variance.

> Conclusion #8: The benefits of CSR experience are apparently not evenly distributed among student groups. African American (Black) students showed stronger gains in achievement associated with small class experience than did other ethnic groups. There is weaker evidence that poor students and children not certified for special education may benefit slightly more from participation in reduced size classes than to those who are not poor or are certified for special education.

Again, these findings represent the marginal contributions of CSR, after controlling for the other factors that influence student achievement. Only in the case of the African American students do inter-group differences reach the level of statistical reliability needed to be confident that the differences found in this study sample would be confirmed in further tests.

> Conclusion #9: Because class size reduction is so deeply entangled with student, school and teacher variables, it virtually impossible to fully disentangle the various factors influencing achievement with the usual *post hoc* exploratory data analysis.

Adequate assessment of the influence of CSR on student achievement will require a convincing conceptual framework capable of directing attention to the specific mechanisms by which CSR is expected to raise student performance. Absent a compelling theory of the mechanisms of performance improvement, it is impossible to know with any degree of certainty which of the very powerful factors examined in this report need to be controlled through planned variation, randomized implementation, or statistical methods when interpreting the data.

Recommendations for Policy Action

The five policy recommendations representing logical extensions of the data analyses presented in this report include:

> Recommendation #1: The most obvious implication of this study is that California would be well advised to "stay the course" with class size until its full effect can be analyzed and documented.

Initial implementation has almost certainly been sufficiently disruptive of school operations that the data analyzed here do not tell the whole story of what can be expected from class size reduction. Until we are able to see how much the academic performance of California's fourth graders can be improved by up to four years of smaller class size experience, it is not appropriate to assert that we really know that CSR does or does not improve student achievement. As Murnane and Levy (1996) found in Austin, Texas, highly effective schools required four years to see consistent growth in achievement from their simultaneous introduction of class size reduction with other major instructional programs.

> Recommendation #2: Given the evidence of rather limited impact during the first two years of implementation of CSR, it is appropriate to begin now testing whether substantial investments in targeted student intervention programs, or expanded professional development activities might contribute more to student learning than a simple reduction in the number of children assigned to a classroom.

Since school program assignments, year round school track assignments, segregation of student groups within the schools, and teacher education and contract status are all more powerfully correlated with student achievement than CSR, it would seem reasonable that policies and programs be developed on the basis of careful examination of how these factors are influencing student learning and how they might be managed to better capitalize on their benefits.

> Recommendation #3: Support needs to be given to work that establishes appropriate explanatory frameworks for interpreting the relationship between class size and student achievement.

To date, research on the relationship between class size and student achievement has been remarkably devoid of meaningful theory. Exactly why removing some children from a classroom should cause the achievement of those remaining to go up remains largely unexplained, even as it is widely expected to be more true than careful data analysis has been able to support.

From the nature of the policy debates informing the adoption of CSR, and from the approaches taken in most research studies, we can infer that there are four competing theoretical frameworks for explaining how smaller classes might be expected to improve school performance. The first, and most common, framework assumes that CSR will work because it increases the instructional resources available to each child in the school. It is assumed that lowering the number of children in a classroom will mean that each child has more access to the teacher and probably

more physical space. As educators or policy makers realize that CSR may have less impact than initially hoped, they start to focus on whether teachers need to change their instructional practices in order to produce the benefits expected from smaller classes. That is, they begin to hypothesize that additional resources alone will not produce results – changed instructional practices, possible only in smaller classes, are required. This instructional change model sees CSR as an opportunity to improve schooling, but one that will only be realized if teachers adopt instructional practices appropriate to the smaller class context. The research literature is not very clear about exactly what instructional changes are needed, and even less clear about why some teachers are more likely to make the appropriate changes than are others.

A third theoretical framework sees CSR as changing classroom organization rather than resources or instructional techniques. This view hypothesizes that smaller classes raise achievement by creating more homogeneous classroom groups and by reducing the frequency with which teachers have to cope with students' learning problems. The fourth theoretical model extends the idea of CSR impact on classroom organization by proposing that smaller classes become effective through the creation of greater student engagement and motivation. The working hypothesis behind this fourth view is that the effectiveness of the smaller classes springs from their ability to reduce alienation and enhance the development of a cohesive community among students and teachers. From this point of view, smaller classes are expected to be most effective in improving the learning of those students most often disengaged from the learning process. Thus, children who have educational handicaps, who are stressed by poverty, or who have been the victims of racial or ethnic prejudice are most expected to benefit more than those from mainstream, middle class families.

Each of these theoretical models is a reasonable account of why we should expect class size changes to produce changes in student achievement. No doubt, there are other reasonable theories. It is important to develop these theories to the point that their implications for achievement patterns and interactions with the student, classroom and school level variables reviewed in this report can be conceptualized and tested.

Recommendation #4: The educational policy community needs to continue the search for school reform and improvement policies that promise to have achievement effects as large as poverty, home language and student ethnicity.

Quite obviously, class size reduction is not the "silver bullet" needed to offset serious educational challenges facing children from poor, minority or non-English speaking homes. Even the most optimistic projections of the achievement gains to be generated through continued and careful implementation of CSR do not lead us to seriously believe that this policy will solve the pressing problems of low achievement haunting California schools.

Recommendation #5: A serious effort needs to be made to strengthen the ability of education researchers and school professionals to develop data systems capable of supporting analysis of relationships between the implementation of specific educational programs and services and resulting changes in student achievement.

Researchers and school professionals interested in documenting the impact of various programs and policies on student achievement find themselves faced with a continuing and serious problem of data availability and usability. Current educational data systems (such as California's CBEDS and STAR data systems) lack two characteristics that are absolutely essential for documentation of policy effectiveness. First, these data systems typically collect only one of the three elements of a program or policy evaluation. To evaluate any program or policy, basic data on school inputs related to student, classroom and school composition must be linked to information on the actual delivery of educational programs and services. These data must, in turn, be linked to measures of student attainment (or other targeted educational goals). California's CBEDS system provides useful data about student characteristics and the teaching resources made available to them (though the system does not enable anyone to know with any degree of confidence which students had access to what teacher or school resources). The STAR data system provides important, though somewhat limited, data on how well students are achieving academic outcomes. There is no comparable data system recording what instructional programs or practices were used by schools or classroom teachers in their efforts to educate the students, however. Even more problematic is the fact that data on student achievement and the records of resources used in their instruction are stored in ways that do not permit continued monitoring of the success of failure of specific educational programs and services. Typical data collections maintain records for a year at a time without permitting tracking student performance from year to year, or continuing analysis of resource availability or program and service delivery processes.

The Impact of California's Class Size Reduction Initiative on Student Achievement: Detailed Findings from Eight School Districts

Douglas E. Mitchell
Ross E. Mitchell
California Educational Research Cooperative
University of California, Riverside

Introduction

Class size reduction is one of the more prominent features of both federal and state level education policy development in recent years. Across about two dozen states, billions of dollars have been spent to lower class sizes to somewhere between 15 and 22 students in the early elementary grades. Among the 13 state initiatives listed in Table 1, eight limit class sizes for all students while five others (Tennessee, Virginia, Wisconsin, Michigan and South Carolina) target funds on low-income schools or districts. A wide variety of reasons for adopting these expensive policies have been offered, including: improving school safety, lowering teacher work loads, strengthening parent participation, creating a more communal atmosphere and enhancing attention to students with special needs. Whatever other outcomes may be expected, however, raising student academic achievement is universally seen as the most basic goal.

In California, the *Class Size Reduction Program* authorized by Senate Bill 1777 in 1996 represents the most expensive educational reform effort ever undertaken by any state (California Department of Education 1996, 1999). State funds allocated during the first two years of operation amounted to nearly $2 billion ($1,827,862,000) for operations and an additional half-billion ($530,905,000) for facilities. In addition to these funds, substantial additional monies were spent from local school district general funds and through waiver authorizations for year round education (California Department of Education 1999).

Recognizing the educational, fiscal and political importance of this initiative, the California Educational Research Cooperative (CERC) initiated a study of the impact of this initiative on school staffing and student achievement. This report presents initial findings from the student achievement phase of that study. The analysis undertaken here builds on previous work undertaken by CERC staff who have analyzed in detail the research literature on class size (see, Mitchell, Carson and Badarak, 1989) and analyzed achievement data from Tennessee's Project STAR (Mitchell, Beach and Badarak, 1989). The current CERC research is being undertaken in a context of close cooperation with the California CSR evaluation consortium which has been funded by the California Department of Education to undertake a statewide review of CSR impacts (California Department of Education 1999).

Research on the effects of class size has a long and colorful history. Early studies produced decidedly mixed results, with some studies concluding that children actually perform better in larger classes. Two landmark studies have shaped the relationship between research and policy on this issue. The first was a comprehensive meta-analysis undertaken by Glass and Smith (1978) which concluded that there is a modest but consistently positive link between class size

reduction and student achievement on standardized achievement tests. An extensively circulated version of this study, published in Educational Evaluation and Policy Analysis (EEPA) (1979), catalyzed national interest and led directly to policy initiatives in several states. In its final form, the Glass and Smith (Glass, et al. 1982) analysis identified the relationship between class size and academic achievement as a logarithmic function (with progressively larger effects produced as each additional student is removed from a classroom). In their widely read EEPA article, however, these researchers approximated the relationship with two straight-line functions. One had a relatively steep slope, showing rapid gains in achievement as students are removed from classes of 17 or fewer students. The other line, approximating the effects for classes of more than 17 students, had a such a small slope (less than a third of an NCE or percentile point for each student removed from the class) that Glass and Smith concluded that class size changes above 17 students would have little or no impact on overall class attainment. Glass and Smith had their critics, but their work was sufficiently convincing that many education policy makers relied on their conclusions to formulate class size reduction initiatives.

Table 1. Current Major Class Size Reduction (CSR) Initiatives Implemented in the United States of America (USA) Since 1984

State	Year	Grades	Target Population	Extent	Citation
Indiana	1984	K-3	All Children (18:1 K-1, 20:1 2-3)	Statewide	Molnar (1998)
Texas	1984	K-4	All Children (22:1 max)	Statewide	Robelen (1998)
Oklahoma	1985	K-6	All Children (20:1)	Statewide	McKeon (1992)
Tennessee	1985	K-3	High % Low Income Districts (15:1)	Statewide	Molnar (1998)
Nevada	1988	K-3	All Children (15:1)	Statewide	Molnar (1998)
North Carolina	1990	1-3	All Children (15:1)	Burke County	Egelson, et al. (1996)
Utah	1992	K-8	All Children	Statewide	Robelen (1998)
Florida	1995	K-3	All Children (15:1 At-Risk Schools, 20:1 Others)	Statewide	Robelen (1998)
California	1996 1998	K-3 9	All Children (20:1 max) English, Language Arts Students	Statewide	Wexler, et al. (1998)
Virginia	1996	K-3	High % Low Income Schools	Statewide	Robelen (1998)
Wisconsin	1996	K-3	>50% Poverty Districts (15:1)	Statewide	Maeir, et al. (1997)
Michigan	1998	K-3	High % Low Income Districts (17:1, 19:1 max)	Statewide	Robelen (1998)
South Carolina	1999	1-3	High % Low Income Districts	Statewide	Robelen (1998)

One state, Tennessee, took the Glass and Smith findings seriously, but skeptically. Rather than adopting a comprehensive, uniform class size reduction policy, this state initiated the nation's

largest systematic test of the relationships between class size, classroom practices and student achievement. This policy initiative provided the data for a substantial collection of landmark studies of the issue, and seemed to conclusively prove that student achievement can be significantly improved through class size reduction (e.g., Finn and Achilles 1990). Analysis and re-analysis of the Tennessee Project STAR data have left some unanswered questions about the extent and nature of achievement effects (e.g., Mitchell, Beach, and Badarak 1989; Finn, et al. 1989; Mosteller 1995; Finn and Achilles 1998; Krueger 1998; Hedges 1999; Konstantopoulos 1999). These questions, while serious, have left most observers confident that there will certainly be at least some achievement gains for most students if early elementary grade class sizes are kept at or near a maximum student-teacher ratio of 17:1.

Issues in Evaluating the Impact of Class Size Reduction

Five important problems limit the capacity of any research study to produce a reliable estimate of the impact of class size reduction on student achievement. First, class size reduction is always accompanied by a variety of simultaneous changes in school population, education policies, school programs and the professional priorities that guide school practices and student assessments. In California, for example, class size reduction has been accompanied by at least a dozen dramatic shifts, including:

1) Passage of California Proposition 227 which has sharply curtailed bilingual education programs,
2) Adoption of a statewide accountability policy forcing multiple assessments of student achievement and requiring reports on all students not reaching grade-level achievement standards,
3) Implementation of a Beginning Teacher Support and Assessment program creating a two year induction program for new teachers,
4) Changes in the funding model for special education which substantially affects local district costs when children are certified for services,
5) Changing economic conditions that affect unemployment and poverty rates in many districts,
6) Continued immigration and relocation which changes the composition of many school populations,
7) A broad reading initiative aimed at changing the focus and effectiveness of early literacy instruction,
8) Changes in regulations regarding the certification of teachers that have changed both the character and timing of pre-service teacher preparation,
9) Support for development of new instructional technologies aimed at providing students with better access to location-independent and multi-media learning opportunities,
10) Adoption of a new statewide standardized achievement test (the Stanford Achievement Test, version 9) and mandated school level public reporting of achievement test scores,
11) Continued implementation of new textbook and curriculum materials adoption cycles (both language arts and mathematics curriculum frameworks were

changed in the last two years) assuring major changes in the scope, sequence and content of subject matter curricula,

12) Addition of ninth grade class size reduction for specific subjects.

A second problem confounding efforts to evaluate the impact of class size reduction is the "embeddedness" of all student achievement in demographic, classroom, school and district factors that are certain to be confounded with any attempt to measure student performance. Whatever its ultimate impact may prove to be, it is quite certain that other factors affecting student achievement are both large enough to obscure its effects and so dynamic that they cannot be considered either stable or randomly distributed. Among the most prominent demographic factors that are known to have effects large enough to obscure class size effects are: family poverty, ethnicity, home language, inter-school transiency and student gender (e.g., Wang, Haertel, and Walberg 1993). School assignment factors that can be expected interact with achievement measurement include: grade to grade cohort achievement variations, special education placements, language proficiency levels, combination grade class assignments, and the effects of grade-level retention (e.g., Balow and Schwager 1990; Reynolds and Bezruczko 1993; Burns 1996; Entwisle, Alexander, and Olson 1997; Mitchell, Destino, and Karam 1997).

At the level of classroom organization within schools are such factors as: the use of year-round or traditional calendars, the willingness of schools to utilize combination grade classes to manage enrollments, and the extent to which students are segregated by socio-economic status, ethnicity, language fluency levels, student gender or special education category (e.g., Zykowski, et al. 1991; Veenman 1995; Burns and Mason 1998; Mitchell and Mitchell 1999).

Teacher assignments also vary from class to class. Confounded with class size reduction we are likely to find variations in teacher experience, age, contract status, ethnicity, gender and educational attainment (e.g., Alexander, Entwisle, and Thompson 1987; Wright, Horn, and Sanders 1997). Finally, school and district boundaries serve to segregate students by neighborhood, culture, socio-economic background and other factors that are not easily measured (e.g., Entwisle, Alexander, and Olson 1997). All of these factors need to be considered as possible sources of achievement variation before we can confidently conclude that students have benefited significantly from taking instruction in reduced size classes.

A third issue that often confuses efforts to evaluate the impact of class size reduction on student achievement is the question of whether we are interested exclusively in the impact on the ***average attainment of all students*** or want to know the extent to which the policy has changed the ***distribution of learning outcomes*** among various types of students (Mitchell, Beach, and Badarak 1989). If, for example, classroom averages remain relatively constant, but previously failing students are now meeting grade-level standards, would that suffice to justify the expense of this policy? Or, if class averages go up, but low attaining students are no better off than they were before, would that be considered a failure? If class averages go up, but the attainment of students is concentrated on the middle range, so that previously high attaining students are no longer moving ahead as rapidly, would that be considered a failure? In short, what ***patterns*** of classroom attainment are being generated, and how are those patterns to be evaluated?

The fourth factor impacting our ability to evaluate the impact of California's class size reduction initiative is the extent to which classroom, school and district implementation procedures may have interfered with (or possibly enhanced) its impact on student learning (e.g., Illig 1997; Hymon 1997; McRobbie 1998; Bohrnstedt and Parrish 1998; Ogawa and Stine 1998; Wexler, et al. 1998). In California, since local school districts had to implement the policy in a matter of a few months, it was difficult to make needed changes in classroom space and teacher recruitment. Schools of education had no advanced warning, with the result that they prepared no surplus of new teachers to take up the large number of new teaching positions created. Construction companies did not have an opportunity to gear up for the production of new classroom facilities. Even if they did anticipate construction needs, there was no early release of construction funds to prepare classrooms. New teachers, not fully qualified teachers, and teachers transferring to new assignments at the last moment had to start instruction of smaller classes in new spaces. Sometimes such irregular spaces as libraries, multipurpose rooms or computer laboratories were converted for the new classes. A significant number of these problems continue into the second and subsequent years of implementation.

Finally, it may be quite important to consider the timing of small classroom exposure and its evaluation through standardized achievement testing. Results from Tennessee's Project STAR indicate that the major effects of class size reduction are experienced during the kindergarten year, or during the first year a child is exposed to this form of instruction (e.g., Finn and Achilles 1990; Krueger 1998; Hedges 1999). If this is generally true, it may not be possible to measure the effects of class size reduction in settings like California where the small class experiences begin in the first, second or third grade and may not be encountered until the students' second or third year of schooling. Additionally, it is possible, that achievement gains produced during an initial exposure to small classes will not be sustained over time. Careful attention to this issue is required before the job of evaluation can be considered complete.

Framing a Class Size Reduction Evaluation Study

Over the long run, if we discount the impact of other changes in policy, the most appropriate and defensible evaluation of California's class size reduction initiative will involve close study of what happens to academic achievement during the students' fourth grade year. Beginning in 1998, some fourth graders were exposed to small class experiences for one or more years. With each succeeding year, a larger proportion of each cohort of students is exposed, and increasing numbers of students have multiple years of small class exposure prior to their fourth grade instructional experience. All other things being equal, if class size reduction succeeds in changing either the rate or the level of student achievement, those changes will become evident in fourth grade student assessment scores. Complete analysis of the fourth grade impact of class size reduction will take several years, however, and it is important to provide preliminary estimates of class size effects and of how students are being affected during the first critical years of implementation.

Study Design

This report offers an initial assessment of the educational experiences of 26,126 students in grades 2 through 4 from eight Southern California school districts. The district enrollments

range in size from about 580 to nearly 36,000 and represent a broad cross-section of urban, suburban and rural settings. The student records selected for analysis are those where complete matching of students with teachers could be made and where complete data on student classroom assignments were available. Seven of the districts had both large and small classes at the same grade level, making it possible to closely study the contrast between large and small class experience in a substantial sub-sample of students.. All available records from students in regular classrooms (i.e., not community schools, individual tutorial students, or special education Special Day Class classrooms) in each of the three study grades within each of the participating districts were included in this study. They consist of second, third and fourth graders in 1,174 classrooms in 83 schools.

Dependent Variables. The dependent variables – reading, mathematics and language achievement – were measured using 1998 Normal Curve Equivalent (NCE) scores from version 9 of the Stanford Achievement Test (SAT-9) as mandated by the California Department of Education. In addition to reviewing the impact of California's Class Size Reduction (CSR) initiative, this report examines the effect of student background, classroom context and teacher characteristics on individual achievement levels (i.e., Total Reading, Total Mathematics and Total Language SAT-9 scores).

Independent Variables. The central independent variable of interest in this study is, of course, class size – the number of students assigned to each teacher. We seek to determine the extent to which providing children in kindergarten through grade three with classes that have a maximum of 20 students (rather than the 28 to 32 students typical of California public schools prior to the adoption of CSR) has a positive impact on their learning. Class size is not the only influence on student learning, however. Painstaking, and often quite expensive, efforts to improve public school performance over the past several decades has taught us that student achievement is shaped by a broad range of potent demographic, social and schooling factors – factors that are often very unevenly distributed across classrooms, schools or school districts.

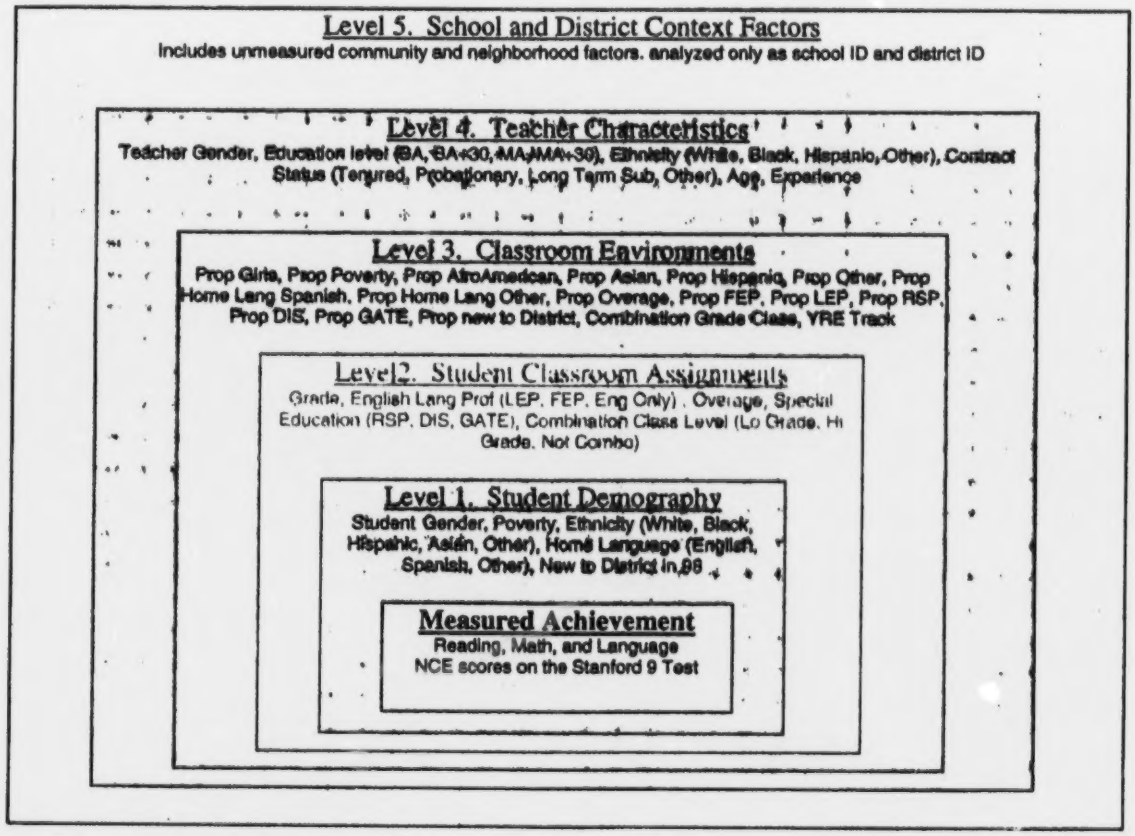
In the study reported here, 19 covariates with potentially powerful impacts on student academic achievement were examined. Fifteen additional variables defining classroom environmental contexts were generated by calculating classroom proportions for each factor level of seven demographic and classroom assignment variables. Taken together, these 34 variables surround and embed student achievement in five distinct contexts or levels. The five levels are depicted in Figure I. At the first level – Student Demography – five factors provide the most fundamental and intractable academic performance influences: gender, family poverty, ethnicity, home language and time of admission to the local district.

At level 2, school organizations begin their influence on student academic opportunities by making class assignments. Five **classroom assignment indicators** – grade level assignment, grade retention resulting in overage students, English language proficiency assessment, special education certification, and the level of placement (upper or lower grade) in combination grade classes – are the most obvious organizational impacts.

Classroom environments constitute the third context level. They are indicated by two variables that operate only at this level – year round education track assignment and whether schools

utilize combination grade classes. Additionally, this study examines fifteen calculated concentration variables that help to define the classroom environment by measuring the classroom proportions of:

Figure I. How Student Achievement is Embedded in Learning Environments


Level 5. School and District Context Factors
Includes unmeasured community and neighborhood factors, analyzed only as school ID and district ID
Level 4. Teacher Characteristics
Teacher Gender, Education level (BA, BA+30, MA, MA+30), Ethnicity (White, Black, Hispanic, Other), Contract Status (Tenured, Probationary, Long Term Sub, Other), Age, Experience
Level 3. Classroom Environments
Prop Girls, Prop Poverty, Prop AfroAmerican, Prop Asian, Prop Hispanic, Prop Other, Prop Home Lang Spanish, Prop Home Lang Other, Prop Overage, Prop FEP, Prop LEP, Prop RSP, Prop DIS, Prop GATE, Prop new to District, Combination Grade Class, YRE Track
Level2. Student Classroom Assignments
Grade, English Lang Prof (LEP, FEP, Eng Only), Overage, Special Education (RSP, DIS, GATE), Combination Class Level (Lo Grade, Hi Grade, Not Combo)
Level 1. Student Demography
Student Gender, Poverty, Ethnicity (White, Black, Hispanic, Asian, Other), Home Language (English, Spanish, Other), New to District in 98
Measured Achievement
Reading, Math, and Language
NCE scores on the Stanford 9 Test

1. a single gender (girls),
2. poverty (children on the National School Lunch Program),
3. overage-for-grade students (15+ months above a September start date for their grade),

Ethnic groups:

4. African-American (black) students
5. Hispanic students
6. Asian students
7. Other non-White students

Different home language groups:

8. Spanish home language speakers
9. Other non-English home language speakers

English language fluency groups:

10. Fluent English Proficient (FEP) students
11. Limited English Proficient (LEP) students

BEST COPY AVAILABLE

Special education category groups:

12. Resource Specialist Program (RSP – educationally at risk) students
13. Designated Instructional Service (DIS – blind, deaf, speech impaired, physically handicapped, etc.) students, and
14. Gifted and Talented Education (GATE) students

Transiency

15. Proportion of students new to the district in the test year

Teacher characteristics comprise the fourth level of influence over student achievement. Confounding the impact of class size we would expect to find significant influence from teacher education and experience levels as well as from teacher gender, ethnicity, age and contract status.

After these variables are all controlled (using statistical controls because experimental procedures are not available), we would still expect unmeasured **school and district level factors** to have a significant influence on student achievement. At this level, we can only examine the extent to which the unmeasured influences associated with student attendance boundaries remain powerful, and to statistically remove them without having any specific explanation as to why they are affecting student test performance.

Table 2. Comparison of Study Sample with California Grades 2 to 4 Student Population

Factor	Level	Sample Mean	California Mean
SAT-9 Achievement (NCE)	Total Reading	42.8	41.0
	Total Mathematics	44.8	42.8
	Total Language	45.0	42.6
Factor	**Level**	**Sample Percent**	**California Percent**
Grade	2	34.1	34.9
	3	33.4	33.0
	4	32.5	32.1
Home Language	Other	4.4	8.8
	Spanish	20.5	30.5
	English	75.1	60.7
English Language Proficiency	English Only	73.9	60.7
	FEP	9.2	8.7
	LEP	16.8	30.6
NSLP Participation	No	57.6	44.5
	Yes	42.4	55.5
Student Ethnicity	Other	2.8	3.6
	Asian	4.9	7.4
	Hispanic	40.0	42.3
	Black	14.4	9.1
	White	37.9	37.6
Student Gender	Male	50.9	50.9
	Female	49.1	49.1
CSR Option 1 in 1996-97 (grades 1-3)	No	49.8	37.6
	Yes	50.2	62.4
CSR Option 1 in 1997-98 (grades 2-4)	No	49.5	44.7
	Yes	50.5	55.3
Teacher's Contract Status	Other	3.4	8.7
	Temporary/L-T Sub	3.5	4.9
	Probationary	28.1	22.9
	Tenure	65.0	63.1
Teacher's Ethnicity	Other	3.7	6.3
	Hispanic	11.8	14.6
	Black	7.0	4.8
	White	77.5	74.1
Teacher's Education Level	BA	20.3	21.1
	BA+30	56.6	55.3
	MA	5.0	11.0
	MA+30	18.0	12.2
Teacher's Gender	Male	17.3	14.0
	Female	82.7	86.0
School's Attendance Calendar	Traditional	51.5	71.7
	YRE 3-Track	2.4	28.3
	YRE 4-Track	46.1	

Describing the Sample

Table 2 presents a statistical comparison of the 26,126 students in our sample with the 1,381,229 California

students in grades 2 through 4. As shown at the top of Table 2, the two groups are closely aligned on overall achievement in reading math and language, while the sample is generally representative of California's total school population, there are half a dozen places where the sample deviates substantially from the overall statewide population. For example, our study sample has more English home language students than the overall state population, with commensurately fewer Spanish home language students. Despite the high number of students from low-income homes (NSLP eligible), the California proportion statewide is yet higher. The sample also has about 20 percent more probationary teachers than the state population, matched by a reduction in the number of teachers on "other" and "temporary" contracts. Similarly, there are more teachers with 30 semester hours beyond the master's degree, matched by a reduction in those holding just the master's degree. The sample also has nearly 20 percent more of its students attending year-round calendar schools than the state population.

Table 3 presents some descriptive statistics for the sample on variables for which statewide population parameters were not available at the time this report was prepared. About 14 percent of the sample students are in combination grade classes. Nearly one out of every eight students was new to the district where they were tested in 1998. Among year-round education tracks, Track C and Track D are the preferred ones. Together they enroll 18 percent more students than Tracks A and B. In our sample, there are only two year-round schools on 3-track attendance calendars, and one of them has a schedule not matched to three of the four tracks in 4-Track calendar schools (1.22 percent of the total sample).

Table 3. Sample Percentages on Factors for which Statewide Comparisons are Not Available

Factor	Level	Sample Percent
Grade in Combo Class	Low	6.59
	High	7.61
	Not	85.80
New to District in 1997-98	No	87.32
	Yes	12.68
Overage for Grade (15+ mos.)	No	96.01
	Yes	3.99
Special Education/GATE	Not	88.51
	RSP	3.58
	DIS	2.17
	GATE	5.74
Enrolled in Combo Class	No	85.80
	Yes	14.20
Attendance Calendar	Traditional	51.49
	YRE "A" 4-Track	10.62
	YRE "B" 4-Track	10.98
	YRE "C" 4-Track	12.88
	YRE "D" 4-Track	12.81
	YRE "A" 3-Track	0.42
	YRE "B" 3-Track	0.38
	YRE "C" 3-Track	0.42

Findings

Investigation of the relationship between CSR and student achievement is undertaken in three steps. First, we examine the extent to which implementation of CSR resulted in the creation of large and small classes that differed systematically with regard to student demographics, classroom contexts or teacher characteristics. Having documented that implementation substantially biased student experiences, we set about to examine the extent to which contextual

factors predict student achievement levels and to examine the impact of these biasing factors on the extent to which class size differences (rather than demographic or contextual factors) are responsible for measured differences in the achievement of students with large and small class experiences. The third step in our study of achievement is to examine the extent to which students with different demographic and context factors derive different benefits from their experiences with reduced size classes. Specifically, we examine whether students with different levels of academic attainment, ethnic background and socio-economic status have different outcomes from their CSR experiences.

Step 1: Discriminating Small from Large Class Experience.

The first step in determining whether CSR contributes to student achievement is to ascertain whether large and small classes are, in other respects, similar. If the two class size settings are systematically different along dimensions that also influence student academic progress, these differences will obscure the true impact of the CSR initiative by inflating (or depressing) student academic performance in ways that add to (or detract from) any marginal difference being generated by changing the number of students in the classroom. Of course, some large and small class differences are expected to ***result from*** the class reduction process itself (for example, the amount of time given to individual students might go up, or the overall noise level in the classroom could go down). These would not be viewed as confounding variables, however. They represent the ***mechanisms*** through which CSR operates to change student performance. The variables of immediate interest are those that potentially confound the CSR effect by directly influencing student performance, independent of the effect of class size. Prior research has, for example, repeatedly found a difference between girls and boys when measured on standardized achievement tests (e.g., Entwisle, et al. 1997). Thus, if one gender group is substantially over represented in either the large or small classes, we would expect this gender imbalance to bias the measurement of large and small class achievement levels, confusing our evaluation of the effects of CSR unless steps are taken to remove this bias from the test score data.

Variables available for this study allow for a fairly detailed analysis of the extent to which California's CSR implementation process created systematic bias in the composition of newly created small classes or inequalities in the teaching resources allocated to them. Tables 4a and 4b, and Figure II display the results of a Multiple Discriminant Analysis (MDA) applied to the 26,126 students in our study sample. MDA is the most appropriate statistical procedure for determining whether several groups are statistically different along a number of simultaneous dimensions. For MDA analysis, the sample was divided into four groups: 1) the 3,047 students with small class experience only in the 1996-97 school year, 2) the 2,480 students with small class experience only in 1997-98, 3) the 8,923 students with small class experience in both years, and 4) the 8,435 students who were in large classes in both years. The remaining 3,241 students were dropped from this analysis because they were missing data on one or more of the variables under study.

Table 4a shows the variables tested, grouped according to four nested levels: student demography, classroom assignment, classroom environment and teacher characteristics. The right column of the table reports the univariate tests of significant difference between the four classroom experience groups (that is, tests of whether the four different class size groups differ

Table 4a. Variables Used to Test Class Size Group Differences

Variables	Univariate Probability that Class Size Groups Differ
Student Demographic Variables	
Poverty	.000
Student Ethnicity:-	
Afro-American	.000
Asian	.000
Hispanic	.000
Student Gender	*.170*
Home Language:-	
Spanish	.000
English	.000
New to District in 98	.000
Classroom Assignment Variables	
Overage for grade	.001
English Language Proficiency:-	
LEP	.000
FEP	.000
Special Education Classification:-	
DIS	.000
RSP	.000
GATE	.000
Classroom Environment Variables	
Classroom demography:-	
Proportion Girls	.000
Proportion Poor	.000
Ethnic Composition:-	
Proportion Afro-American	.000
Proportion Asian	.000
Proportion Hispanic	.000
Language Learner Density:-	
Proportion Spanish Speakers	.000
Proportion English Speakers	.000
Classroom Assignment Composition:-	
Proportion Overage	.000
Special Education Composition:-	
Proportion DIS	.000
Proportion RSP	.000
Proportion GATE	.000
Language Proficiency Composition:-	
Proportion LEP	.000
Proportion FEP	.000
School type:-	
Year Round School in 98	.000
Transiency	
Proportion New to District in 98	.000
Teacher Characteristic Variables	
Teacher Contract Status:-	
Probationary Teacher	.000
Tenured Teacher	.000
Teacher gender	.000
Teacher Experience in years	.000
Teacher Education:-	
Teacher has BA + 30 units	.002
Teacher has MA or higher	.013
Teacher Ethnicity:-	
Afro-American Teacher	.000
Hispanic Teacher	.000
Other Ethnicity Teacher	.000

on each variable without considering the extent of correlation or redundancy among the variables). All but one of the variables (student gender) were found to have statistically significant different values across the four class size groupings. Thus, there is a very strong *prima fascia* case for believing that, at least during the first two years of implementation, reduced size class experience was far from evenly distributed across students of various backgrounds or classrooms with differing compositions.

Table 4b summarizes the multivariate, multiple discriminant analysis, findings regarding class size implementation group differences. As shown near the top of the table, there are three significant multiple discriminant functions separating the four class size experience groups. Taken together, these discriminant functions account for 35.2 percent of the variations in class composition, which means that knowing the values of these variables for a student increases by more than one-third the probability that we could accurately predict his/her small and large class experience. This table provides the evidence to support three important conclusions.

1. The composition and context differences separating large and small class experiences for California school children were quite substantial during the first two years of CSR implementation.

While 35.2 percent of the variance in group membership predicted by this discriminant analysis does not allow anything close to perfect prediction of CSR experience, it does indicate a very substantial bias in the composition of large and small classes. On virtually every parameter for which we have data, students in small classes differed significantly from those in the larger classes.

2. The most important differences between the large and small classes include:

    a) Classroom ethnicity – particularly the proportion of Afro-American students
    b) Poverty – the proportion of each class drawn from poor families
    c) Transiency – the proportion of children in a classroom new to the district
    d) YRE – whether the classes are located in Year Round calendar schools
    e) Special education designations – particularly the proportion of the class which is GATE or RSP certified
    f) Teacher gender – whether the class is taught by a man or a woman

Twenty-one other variables add small, but statistically significant increments to the overall contrast in composition, context and teacher characteristic differences between large and small classes. That is, multivariate analysis identifies a total of 26 variables on which the classrooms experiencing different CSR implementation patterns during the first two years differ in statistically significant ways.

In the multivariate analysis, the variables that do not add statistically significant increments to distinguishing among the CSR classroom types are those measuring *individual* student characteristics, rather than classroom composition. These include individual characteristics related to: ethnicity, home language, poverty status, special education designation, student

Table 4b. Multiple Discriminant Analysis of the Extent to which CSR Implementation was Biased by Student Demographics, Classroom Assignments, Classroom Contexts, or Teacher Characteristics

Coefficients appearing in **bold** are more than 50% of the maximum coefficient for each respective function; coefficients appearing in *italics* are the maximum coefficient for each respective variable

	Discriminant Structure Matrix Coefficients		
Variables	Function 1	Function 2	Function 3
Classroom Proportion Black	***-0.583***	-0.096	-0.070
Classroom Proportion Low Income Status (NSLP)	***-0.423***	-0.236	0.120
Student New to District in 1997-98	*0.267*	0.000	0.250
Classroom Proportion FEP	*-0.257*	-0.083	0.145
Low Income Status (NSLP) Student	*-0.245*	-0.132	0.080
Black Student	*-0.227*	-0.043	-0.025
Classroom Proportion Hispanic	*-0.152*	-0.074	0.019
FEP Student	*-0.134*	-0.023	0.036
Teacher with Tenure Contract	*-0.122*	-0.044	-0.114
Teacher with Probationary Contract	*0.063*	0.022	-0.016
Hispanic Student	*-0.058*	-0.037	0.018
Student Overage 15+ Months	*0.039*	-0.026	-0.017
Teacher with Master's Degree	*-0.038*	0.019	-0.020
Classroom Proportion GATE	0.067	***-0.561***	0.015
Female Teacher	0.052	***0.456***	-0.145
Classroom Proportion RSP	0.189	***-0.327***	-0.251
Black Teacher	-0.144	*-0.280*	0.122
GATE Student	0.021	*-0.263*	0.003
Classroom Proportion Asian	0.103	*0.224*	0.150
RSP Student	0.010	*-0.158*	0.002
Combination Grade Class	0.009	*-0.138*	0.026
Asian Student	0.063	*0.101*	0.066
Teacher with Bachelor's Degree + 30 Sem. Hrs.	-0.014	*-0.068*	0.020
DIS Student	0.003	*0.048*	0.039
YRE School (YRE Track Classroom)	**-0.424**	-0.132	***-0.541***
Classroom Proportion New to District	-0.186	0.101	***0.522***
Classroom Proportion Female	0.061	-0.106	*-0.231*
Classroom Proportion English Home Language	0.079	-0.021	*-0.198*
Teacher Total Years in Position in 1998	-0.110	-0.028	*-0.191*
Classroom Proportion LEP	0.083	0.051	*0.182*
Classroom Proportion Spanish Home Language	-0.091	0.007	*0.165*
LEP Student	0.137	0.037	*0.156*
Other Ethnicity Teacher	0.058	-0.014	*0.155*
English Home Language Student	-0.011	-0.019	*-0.143*
Classroom Proportion DIS	-0.020	0.131	*0.137*
Spanish Home Language Student	-0.017	0.007	*0.127*
Classroom Proportion Overage	0.090	-0.051	*-0.103*
Hispanic Teacher	0.015	0.049	*0.094*
Female Student	0.004	-0.011	*-0.047*
Squared Canonical Correlation (R^2)	0.216	0.112	0.069
Total R^2 = 0.352	$p<.0005$		

	Function Evaluated at CSR Group Centroids		
CSR Class Experience Type	Function 1	Function 2	Function 3
CSR Only 1996-97	-1.13	-0.42	-0.18
CSR Only 1997-98	0.71	-0.07	-0.68
CSR in Both Years	-0.13	0.43	0.05
No CSR Experience	0.34	-0.28	0.22

BEST COPY AVAILABLE

BEST COPY AVAILABLE

gender and being overage for grade. These variables, while differing substantially across class types are less powerful than the composition variables constructed by calculating the proportion of students with each of these characteristics found in the different CSR class types.

Two notes are important at this point in our analysis. First, in addition to the substantial direct contributions to differentiating among CSR implementation settings made by the variables reported in Table 4a, there may also be significant interaction effects created by the their combined impact. Our focus in this analysis, however, is to determine if it is important to consider differences in the composition of large and small classes in order to determine the true effect of CSR implementation on student achievement. Even without considering potential interactions among the available variables, the answer to this question is clearly a resounding "Yes!," making it necessary to look closely at these confounding variables whenever we try to estimate the impact of CSR on overall student achievement.

The second important note is that we have no data on numerous other factors that might serve to distinguish the composition of the large and small classes. For example, we do not know how many of the experienced teachers had experience at the specific grade levels of their current classroom assignments. And we know nothing of whether the teachers in small classes had any specific training in techniques that might be appropriate for this class size condition.

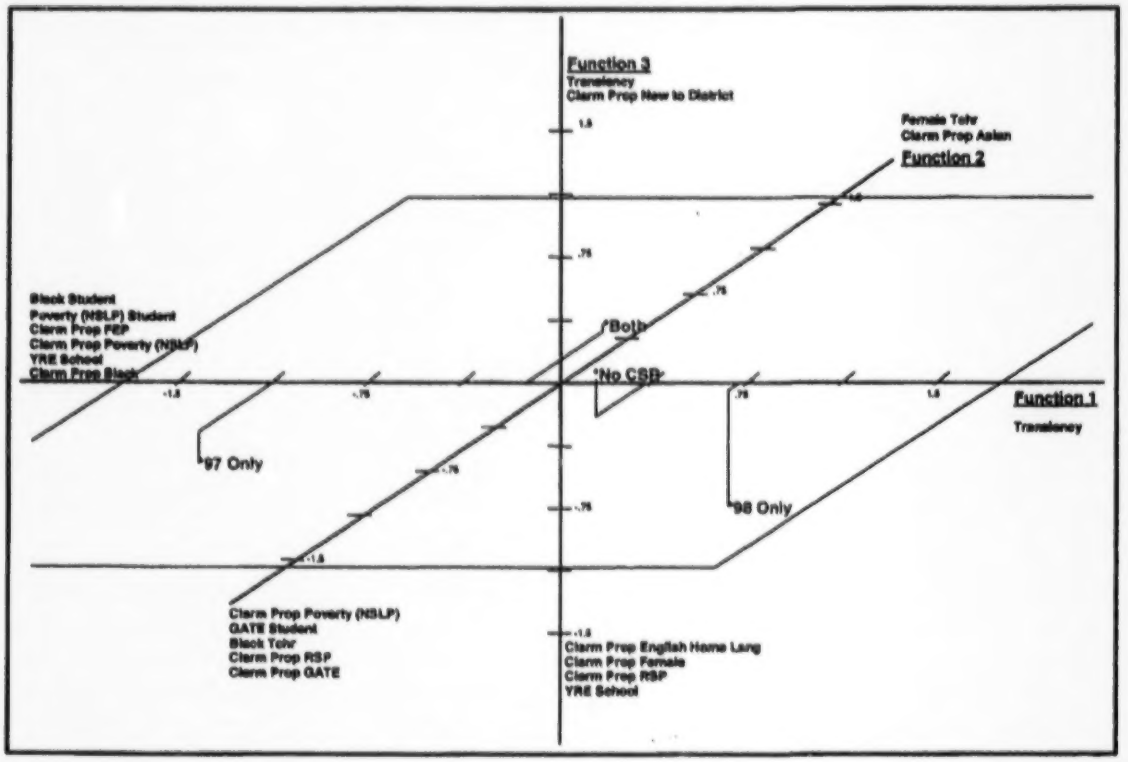
The "bottom line" is simply this: ***the differences among the four different CSR implementation conditions (small in 96-97 only, small in 97-98 only, small in both years, and large in both years) are very large and involve a broad array of class composition and context variables.***

3. The most striking contrast in CSR implementation settings is between those where classes were small only in 1996-97 and those where the classes were small only in 1997-98.

The group centroids for the four different types of CSR experience, shown in the bottom section of Table 4b, identify the center of density for each class type. By plotting these centroids in 3-dimensional space, as in Figure II, it is possible to see just how the groups differ. Note, for example, that the classes composed of students who experienced small classes only in 1996-97 (labeled **'97 Only'** on the figure) are located near the left end of the axis labeled "Function #1." Placement here indicates that this group of students was much more likely to be in classes that were: higher in the proportion of poverty, African American(Black) and Limited English Proficiency students and to be located in year round calendar schools. By contrast, those experiencing small classes only in 1997-98 (labeled **'98 Only'** toward the right side of the figure) were less likely to have these characteristics and more likely to be composed of students new to the district in that year. By observing the variables identified at the ends of the axes in Figure II, and following the placement of the group centroids it can be seen that students with no small class experience, and those with experience only in 1996-97, are more likely to have a high proportion of GATE students, and less likely to have female teachers than the other groups (they have the negative centroids on Function #2). Similarly, students with small class in one or the other of the implementation years, but not in both years, are more likely to be in year round calendar schools and to have higher proportions of RSP students in their classes (i.e., these two groups have the negative centroids on the third discriminant function).

There does not appear to be any simple explanation for the marked differences among the four CSR implementation class types. School leaders may have tried to focus CSR on poor, minority and year-round calendar students during the first year to maximize their learning opportunities. And that the composition of the second year implementation groups was, as a consequence, made up disproportionately of more advantaged students. It might also be that it was politically easier to move the poor and minority students during the first hectic months of CSR implementation. What ever the reasons may have been, however, the important point is that the small classes created during the first year were substantially different in composition from those created during the second year. Moreover, the students with small class experience during both years, and those with no small class experience at all were characterized by unique demographic

Figure II. Three Dimensional Plot of MDA Group Centroids Contrasting Four Types of Class Size Reduction Experience



composition, classroom contexts, and teacher characteristics. Thus, we must carefully examine the relationship between these implementation biases and student achievement before we will be able to tell whether CSR itself makes any substantial contribution to academic attainment.

BEST COPY AVAILABLE

Step2:- Documenting the extent to which the CSR Impact on Academic Achievement is Confounded with Demographic, Classroom Context and Teacher Variables.

Tables 5a through 5e display the direct effects of each of the variables identified as possible biasing factors confounding the influence of CSR implementation on student achievement. The variables tested are those utilized in the discriminant analysis described above, with the addition of the four CSR implementation conditions (Small class only in 1996-97, only in 1997-98, both of the first two years, and no experience in reduced size classes). The variables tested are identified on the left side of each table, with the impacts on Stanford 9 reading, mathematics and language scores shown in the three columns toward the right side of the tables. The far right column of each table provides an overall summary of the magnitude of each variable's main effect impact on achievement (ranging from "Negligible" to "Very Large"). At the top of the reading, math and language columns are the grand means and standard deviations for each test for all of the 26,126 students in the study sample. For each of the categorical variables (e.g., poverty, ethnicity, home language) the numbers shown in the body of each table are the "estimated marginal means" generated by a General Linear Models statistical procedure (using the *Statistical Package for the Social Sciences 9.0*®). For each factor, a reference level is identified (shown in the column labeled "Reference Category") and the marginal mean differences are shown as the NCE score difference between each other factor level and the reference category (i.e., factor level minus reference category). Thus, the numbers shown in each row for these variables represent the extent to which students who fit into the category specified in the column labeled "Level" gain an advantage (positive numbers) or suffer a penalty (negative numbers) relative to the reference category. In Table 5a, the CSR reference category is "No Small Class Experience," with the numbers in the three rows for each CSR implementation condition representing the gain (positive numbers) or loss (negative numbers) associated with being placed in a small class in either the first year, the second year or both the first two years of implementation.

For each continuous variable examined in Tables 5d and 5e (e.g., the proportion of students in poverty or the proportion of GATE students in a classroom), the numbers in the table are the "unstandardized regression coefficients" also generated by the General Linear Model statistical procedure. For all the continuous variables except teacher experience and age, the unstandardized regression coefficients indicate how much achievement test scores change as the classroom proportion of a given characteristic moves from zero (none of the covariate present in a classroom) to 1.0 (the classroom consists of 100 percent of the measured covariate). For teacher age and experience, the unstandardized regression coefficients represent the amount of change in a student test scores (measured in NCE points) resulting from a one-year increase in teacher age or experience.

The variables analyzed in Tables 5a through 5e are separated into five successive analyses, representing the first five levels described in Figure I, above: a) raw CSR impacts, not considering confounding covariates; b) student demographics; c) classroom assignments, d) classroom environments and e) teacher characteristics. (Since CSR is a classroom level policy, the school and district level factors were examined only to determine how much additional achievement variance they might account for, not to provide statistical control of variables that might influence the extent of CSR impact.

At the bottom of Tables 5a through 5e, the total percent of variance explained by the variables included in each level of analysis is reported. At the bottom of Table 5e, we report how much additional variance in student achievement is explained when school and district characteristics are added to the model to account for neighborhood and regional student segregation. In the column at the right side of each table, an effect is labeled "Very Large" if the marginal means are separated by an amount equal to or greater than one-half of a standard deviation for the overall sample (this amounts to approximately one grade level difference on a grade equivalent scale). Effects are deemed "Large" if they reflect marginal mean differences of about one-quarter to one-half a standard deviation. They are labeled "Moderate" if they reflect differences in the .15 to .25 standard deviation differences (about the size of the reported achievement differences in Tennessee's Project STAR experiment). They are labeled "Small" if they are less than .15 standard deviations and "Negligible" if they are not statistically reliable.

Table 5a. Differences in Mean Achievement Related to CSR Implementation for 26,126 Students in Eight School Districts, in Grades 2 through 4

Numbers are mean differences for Stanford Achievement Test, NCE Total Reading, Math and Language Scores (Cell entries are Factor Level NCE's minus the Reference Category identified for each Factor, Unstandardized regression coefficients for continuous variables)

Statistically Significant Values are: Bold p < .001, *Italic* p < .01, *Italic* p < .05

			Grand Mean:-	42.80	44.78	44.98	
			Std. Deviation:-	21.07	22.01	21.47	Overall
Uncontrolled		Reference Category	Level	Reading	Math	Language	Effect Size
	Class Size Reduction	No CSR Experience	1996-97 Only	*-4.9*	*-5.2*	*-3.7*	Moderate
			1997-98 Only	-0.3	*1.8*	-0.6	Negligible
			2 Years CSR	0.1	*1.8*	-0.2	Negligible
Pct of Total Variance Explained by Class Size Reduction				0.6%	1.0%	0.3%	

The apparent CSR impact when other variables are not considered. Table 4a provides our first look at student achievement with and without the experience of CSR generated smaller classes. We begin by noting that the overall means on the three tests analyzed here (reading, math and language) fall between 42 and 45 NCE points – substantially below the Stanford 9's normed mean score of 50 points. Since grade equivalent differences range from 10 to 20 NCE points, depending on the grade level tested, this means that the children in this sample scored between a third and two-thirds of a grade level below average on the norms established for this test. While this places our sample significantly below grade-level, the students in our sample are reasonably typical of California students in grades two through four whose statewide average NCE scores were about: 41 in reading, 43 in math and 43 in language.

As shown in the middle part of Table 5a, initial estimates of the overall impact of California's class size reduction initiative on student achievement are rather disappointing. Students who experienced small classes only during the initial year of implementation (1996-97) diverge most sharply from those who had no small class experience, ***but this difference is in the negative direction – in reading, students attending small classes in 1996-97 but returning to large classes in 1997-98 were 4.86 NCE points <u>below</u> those with no small class experience. In mathematics the picture is slightly worse with the first-year small class participants scoring 5.18 points below the students with no small class experience, in language the difference is***

BEST COPY AVAILABLE

slightly smaller, but still a negative 3.75 points. Students who experienced smaller classes only in the second year of implementation, and those who were in the small classes in both years fared a bit better. The difference is slightly negative or near zero in reading and language, but in mathematics these two groups outscored the no-small-class-experience group by nearly two NCE points (1.84 and 1.79, respectively).

As shown at the bottom of Table 5a, while the apparent CSR effects are statistically significant, the account for a mere 0.6 percent of the total variance in student achievement – too small to be considered a major factor in shaping student academic performance.

Table 5b examines the extent to which student demographic factors influence student achievement and alter the picture of how CSR might be affecting student academic performance. There are two important points to be observed in this table. ***First, each of the student demographic factors affect student achievement much more powerfully than does CSR experience.*** Being poor, for example, lowers students ability to score well on the Stanford 9 test by at least 11 or 12 NCE points. This is more than half a standard deviation and represents test performances nearly a full grade level behind children who do not qualify for free or reduced price lunches. Having a non-English home language also interferes with a student's ability to score well on the Stanford 9 test battery. Children from Spanish speaking homes score from 10 to 13 NCE points below their English home language peers – again nearly a full grade level below. Note that these are the *independent* effects of poverty and home language, meaning that poor children from Spanish speaking homes are likely to be doubly jeopardized – scoring more than 20 NCE points or nearly two grade-levels below their non-poor, English home language

Table 5b. Differences in Mean Achievement Related to Demographic and Schooling factors for 26,126 Students in Eight School Districts, in Grades 2 through 4

Numbers are mean differences for Stanford Achievement Test, NCE Total Reading, Math and Language Scores (Cell entries are Factor Level NCE's minus the Reference Category identified for each Factor, Unstandardized regression coefficients for continuous variables)

Statistically Significant Values are: Bold $p < .001$, Italic $p < .01$, Italic $p < .05$

			Grand Mean:-	43.79	44.78	44.96	
			Std. Deviation:-	21.07	22.01	21.47	Overall
Student Demographic Variables		Reference Category	Level	Reading	Math	Language	Effect Size
1	Low Income Home (NSLP)	Not Poor	Poor	-11.30	-12.18	-11.78	Very Large
2	Student Gender	Female	Males	-4.06	-1.49	-6.36	Large
3	Student Ethnicity	White	Asian	4.18	8.82	4.60	
			Black	-4.66	-7.33	-5.01	Large
			Hispanic	-2.16	-2.56	-4.86	
			Other	-1.65	-2.66	-1.10	
4	Home Language	English	Spanish	-12.96	-9.77	-10.86	Very Large
			Other	-3.19	0.29	-2.34	
5	Student New to District	Continuing Student	Mobile Students	-3.98	-3.76	-4.31	Moderate
Cumulative Pct of Total Variance Explained without considering Class Size Reductio				24.4%	22.0%	22.5%	
On Student Demographic Residuals		Reference Category	Level	Reading	Math	Language	
*	Class Size Reduction	No CSR Experience	1996-97 Only	*-1.8*	-1.7	-0.7	Negligible
			1997-98 Only	-1.3	0.6	*-1.8*	Negligible
			2 Years CSR	0.2	*2.0*	-0.1	Small
Pct of Total Residual Variance Explained by Class Size Reduction				0.1%	0.4%	0.1%	

peers. Student ethnicity is another Very Large factor influencing SAT-9 test performance with 9 to 17 NCE points separating the highest and lowest performing ethnic groups on each of the SAT-9 sub-tests. Moreover, even as these important factors are being taken into account, student performance is strongly being affected by student gender (girls outscore boys by one and a half to six NCE points) and transiency (children new to the district lose about 4 NCE points).

BEST COPY AVAILABLE

The second important point underscored by Table 5b is that statistically controlling for the effects of student demographic factors on achievement dramatically alters the picture of what CSR implementation might be doing for student performance. As can be seen from the estimated marginal mean differences for CSR implementation class types (shown toward the bottom of Table 4b), the achievement impacts of the various student demographic factors are significantly confounded with the effects of CSR implementation. As predicted by the discriminant analysis showing substantial differences in the composition of the small and large classes described in a previous section, the apparent negative consequences of experiencing small classes only during the first year of implementation are largely removed by controlling for the demographic composition of these early implementation classes. What appeared to be a 4 or 5 NCE point penalty for early CSR implementation has been reduced to only about a 1.5 NCE point spread between the large classes and the first year CSR classes.

The picture for the second year only CSR classes has also been dramatically affected, however. In this case, controlling for the effects of student demographics lowers the estimate of how experiencing smaller classes in 1997-98 has impacted on student academic performance. For each of the three SAT-9 subtests, removing the student demographic influences on achievement lowered the estimated performance for the second year implementation group by more than 1 full NCE point. For the students who were in small classes in both 1996-97 and 1997-98, removing demographic influences on achievement altered the estimates of SAT-9 performance only slightly.

As noted at the bottom of Table 5b, when student demographics are included with CSR in an analysis of factors influencing student achievement, we are able to account for a respectable 22.0 percent (math) to 24.4 percent (reading) of all the variations in student achievement.

Interaction effects. Included in the statistical models, but not shown on Tables 5a through 5e are very substantial interaction effects among the various student, classroom and teacher variables under study. The two most powerful and important interactions among the demographic variables are those between home language and poverty and between ethnicity and poverty.

The interaction between home language and poverty is shown on Table 6, and depicted graphically in Figure III. Table 6 shows how dramatically these two variables, working in combination threaten academic success. Among the not poor students, Spanish-speaking children still fall more than a grade level behind their peers (the non-poor speakers of other non-English languages pay a much smaller price for their lack of English experience at home). Poor children who do not come from English speaking families suffer an additional grade-level size loss in achievement. That is, poor children from non-English speaking homes (particularly Spanish speaking homes) are nearly two full grade levels below their English speaking, not poor peers (at least when tested in English).

As the bargraph in Figure III shows, the interaction between home language and poverty is such that the achievement of Spanish speaking poor children, is closer to their English speaking peers than is that among the not poor children. Depending on which subtest is being considered, the

gap in academic performance among the not poor children is from 40 to 100 percent larger than among the poor children.

Figure III. Interaction of Student's Home Language with Home Income Status (NSLP) on SAT-9 Total Reading, Mathematics and Language Achievment

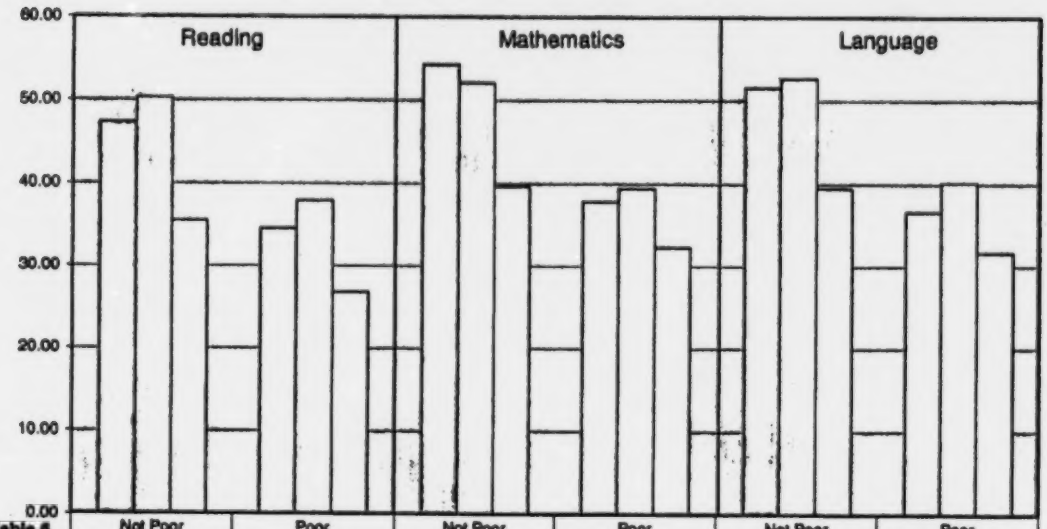


Table 6	Not Poor	Poor	Not Poor	Poor	Not Poor	Poor
Other	47.31	34.50	54.27	37.83	51.58	36.58
English	50.30	37.89	52.08	39.43	52.68	40.17
Spanish	35.47	26.80	39.66	32.31	39.46	31.68

Strong interactions among student demographic factors are even more dramatically seen in Table 7 and Figure IV, where the statistically significant interaction between poverty and ethnicity is reported. For all ethnic groups, poor children fall 12 to 14 NCE points behind their not poor peers. More important, ***among students in poverty, SAT-9 performance does not depend very strongly on ethnicity***. This is not at all the case ***among not poor students*** where ***academic performance is very closely related to ethnicity***. In reading, for example, the average deviation among the poor ethnic groups is less than 1 NCE point, but among the not poor ethnic groups the spread is nearly five times as large. While the contrast is not as large in mathematics and language, the not poor ethnic groups still vary two to three times as much as the poverty groups. In effect, this table and graph show that ***the negative effects of poverty fall fairly equitably on all ethnic groups, but the advantages gained by moving above the poverty line are far from equally shared.*** Higher income Asians and Whites show much more substantial achievement gains than African Americans (Blacks) and Hispanics.

Figure IV. Interaction of Student's Ethnicity with Home Income Status (NSLP) on SAT-9 Total Reading, Mathematics and Language Achievment

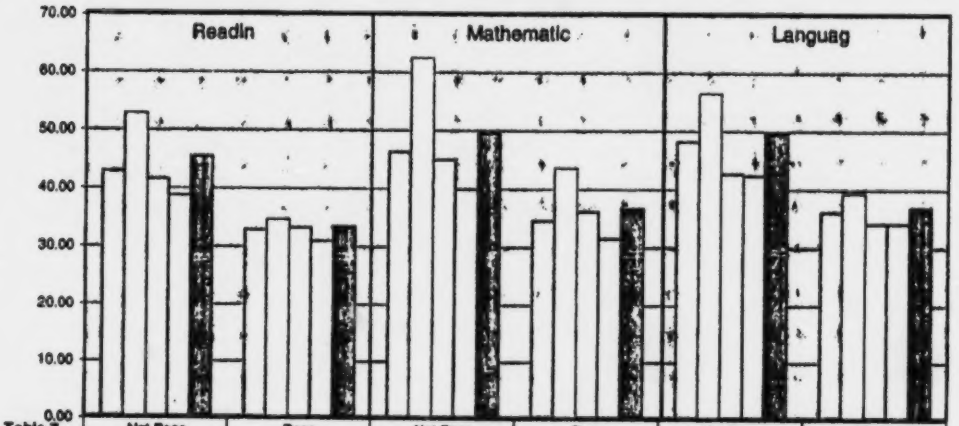


Table 7	Not Poor	Poor	Not Poor	Poor	Not Poor	Poor
Other	42.92	32.87	46.33	34.58	48.25	36.15
Asian	52.86	34.65	62.51	43.84	56.44	39.36
Hispanic	41.57	33.27	45.02	36.12	42.70	34.18
Black	38.83	31.01	39.95	31.60	42.37	34.20
White	45.63	33.52	49.54	36.70	49.77	36.82

For the SAT-9 reading subtest, the interaction between ethnicity and home language is shown on Table 8 and plotted in Figure V. Cross-classification groups with fewer than 200 students represent only about 10 classrooms and are not shown (or plotted on the graph). Groups this small cannot be interpreted as revealing any substantial effects. Note that the no language by ethnicity subgroups show consistently higher achievement than English speaking White students in reading. The English speaking Asians do slightly better in mathematics and about the same on the language test, but do a small amount less well in reading.

BEST COPY AVAILABLE

Figure V. Interaction of Student's Ethnicity with Home Language on SAT-9 Total Reading Achievment

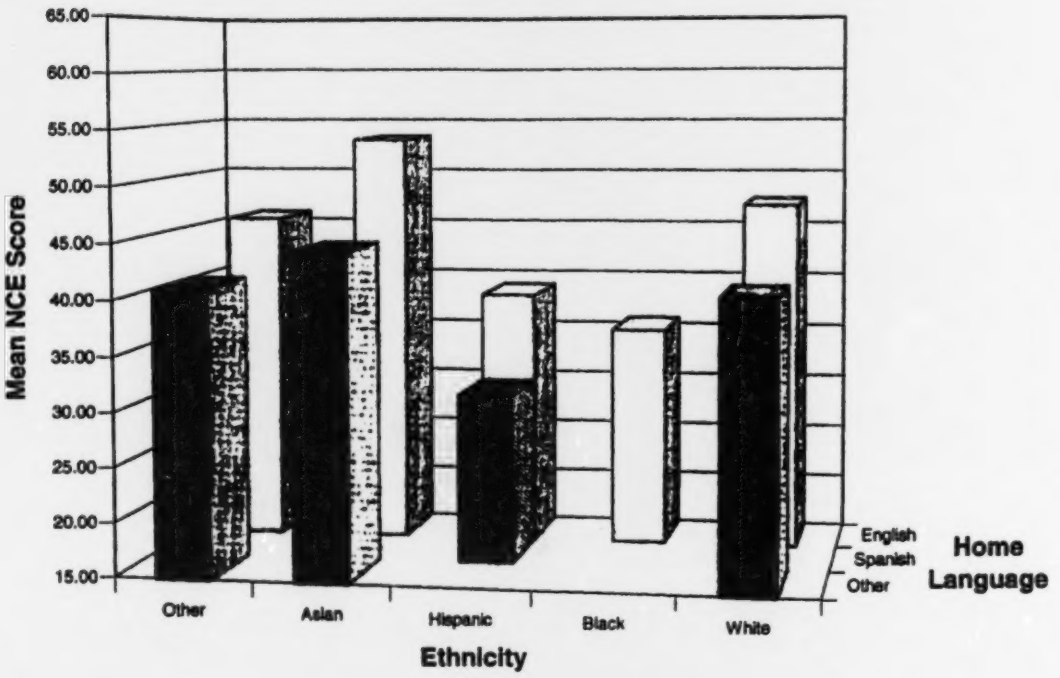


Table 8	Other	Asian	Hispanic	Black	White
■ Other	41.05	44.40			40.90
■ Spanish			30.53		
□ English	45.72	53.30	38.68	35.47	47.30

How Classroom Assignment Variables Influence Achievement.

Table 5c presents the main effects of five classroom assignment variables on reading, math and language test scores. Not surprisingly, the most powerful classroom assignment variable is the differentiation among the special education categories. Special education students are classified into three broad categories. Resource Specialist Program (RSP) for students suffering learning handicaps; Designated Instructional Services (DIS) for students who are have impaired vision, hearing, speech or other physical limitations, and Gifted and Talented Education (GATE) for children identified as functioning at high levels. (Severely challenged students assigned to special day classes or other special education settings were eliminated from the study). Since special education categories are created to differentiate educational services on the basis of children's classroom performance, we would be surprised if there were not Very Large inter group differences on this factor. On average, GATE children score more than 22 NCE points (about two grade levels) above their peers; RSP students more than one grade level behind.

Classification of students according to their English language proficiency is also used by the schools to develop specialized program opportunities. This student assignment factor is Moderately aligned with achievement. Fluent English Proficient (FEP) students show achievement scores significantly above others largely because school districts require a minimum academic achievement level (usually above the 30th percentile) before students are eligible for redesignation from limited to fluent speakers. Limited English Proficient (LEP) students score modestly below English only students and substantially below their FEP peers.

Table 5c. Differences in Mean Achievement Related to CSR Implementation for 26,126 Students in Eight School Districts, in Grades 2 through 4

Numbers are mean differences for Stanford Achievement Test, NCE Total Reading, Math and Language Scores (Cell entries are Factor Level NCE's minus the Reference Category identified for each Factor, Unstandardized regression coefficients for continuous variables)

Statistically Significant Values are: Bold p < .001, *Italic* p < .01, *Italic* p < .05

			Grand Mean:-	42.80	44.78	44.96	
			Std. Deviation:-	21.07	22.01	21.47	Overall
Classroom Assignment Variables		Reference Category	Level	Reading	Math	Language	Effect Size
6	Grade	Second Grade	Third Grade	*-2.24*	0.31	-0.03	Small
			Fourth Grade	-0.92	-2.22	1.95	
7	Special Education	Regular Students	RSP Students	**-12.84**	**-13.20**	**-12.12**	
			DIS Students	*-6.51*	-2.39	*-6.14*	Very Large
			GATE Students	**23.08**	**22.92**	**22.16**	
8	Language Proficiency	English Only	FEP	**4.67**	**4.88**	*3.19*	Moderate
			LEP	-2.08	-1.25	-2.16	
9	Overage for Grade	Not Overage	Overage 15+ mos.	-0.14	-0.61	-1.01	Negligible
10	Combination Grade Class	Single Grade Classes	Lower Grade Combo	0.33	0.90	0.22	Small
			Upper Grade Combo	-2.83	-2.91	-3.20	
Cumulative Pct of Total Variance Explained without considering Class Size Reduction				34.4%	30.3%	31.2%	
On Classroom Assignment Residuals		Reference Category	Level	Reading	Math	Language	
*	Class Size Reduction	No CSR Experience	1996-97 Only	*-1.8*	*-2.3*	-0.7	Small
			1997-98 Only	0.2	0.2	0.3	Negligible
			2 Years CSR	*0.9*	*0.8*	*1.2*	Negligible
Pct of Total Residual Variance Explained by Class Size Reduction				0.3%	0.3%	0.2%	

When controlled for other factors, there is a Small decline in test performance across the three grades in this sample – the second graders outperformed the third graders in reading and the fourth graders in mathematics by more than 2 NCE points. There is also a Small relationship between student assignment to the upper or lower grades in combination classes and their SAT-9 scores. Students in the upper grades of combination grade classes generally score two and a half

BEST COPY AVAILABLE

or three points below students in single grade classes; those in the lower grade in combination grade classes generally do a half-point to a point better.

A new CSR analysis, revised now to control for classroom assignment variables, is shown at the bottom of Table 5c. This analysis reveals that the picture of how small class experience might be affecting student achievement changes again when classroom assignment biases are removed from the data. The order of magnitude of the differences between students with the different types of small class experience and those with none is still quite small – about the same as we saw when only student demographic biases were removed. After controlling for classroom assignment bias, however, the "winners" and "losers" look quite different. There is no apparent pattern to the changes. Students with reduced class size only in the first year appear to drop in reading and math, but not language. Students with experience in the second year only, or in both years, appear to do slightly better in reading and language, but not math. Perhaps the most important point here is that ***the changes in apparent CSR impact are about the same size as the total effect itself***, indicating that the biases in the allocation of reduced size class opportunities are accounting for as much difference in student achievement as is participation in the smaller classes.

Shown just above the renewed CSR analysis, we note that the classroom assignment variables add 10 to 13 percent to our overall ability to explain student academic performance on the SAT-9 test, bringing the total explained variance to: 34.4% in reading, 30.3% in mathematics and 31.2% in language. And, as shown at the very bottom of the table, CSR implementation is accounting for only about three-tenths of one percent of the SAT-9 scores.

How Classroom Environment Variables Influence Achievement

Table 5d reviews the impact of 17 classroom environment variables on SAT-9 scores. There are Very Large effects on achievement related to the proportions of special education students, overage students and specific ethnic groups within the classrooms. The numbers in these cells of the table are unstandardized regression coefficients, but they indicate that changes substantially in excess of a half a standard deviation can be expected as classroom composition moves from zero to 100 percent of the specified groups. While the "proportion of other ethnicity" (which includes Pacific Islanders, Native Americans, Alaska Natives and others) has the largest regression coefficient on each test, the number of students involved here are too small to be considered reliable (even though the coefficients do reach the level of statistical reliability using the regression technique applied here). Substantively more important, the data in Table 4d reveal that concentrations of Asian and Hispanic students provide a substantial advantage to student learning for those in these classes. This class composition advantage does not apparently accrue to African American students.

High concentrations of special education students within a classroom have common sense impact of lowering achievement when the concentration is of low achievers and raising it for high achievers. What may be a little surprising is the extent to which a concentration of special education students affects class attainment. Beyond the individual achievement impact of a

Table 5d. Differences in Mean Achievement Related to CSR Implementation for 26,126 Students in Eight School Districts, In Grades 2 through 4

Numbers are mean differences for Stanford Achievement Test, NCE Total Reading, Math and Language Scores (Cell entries are Factor Level NCE's minus the Reference Category identified for each Factor, Unstandardized regression coefficients for continuous variables)

Statistically Significant Values are: Bold p < .001, *Italic* p < .01, *Italic p <.05*

			Grand Mean:-	42.80	44.78	44.96	
			Std. Deviation:-	21.07	22.01	21.47	Overall
Classroom Environment Variables		Reference Category	Level	Reading	Math	Language	Effect Size
11	Combination Classes	Single Grade	Combo	*0.96*	0.17	*1.12*	Small
12	YRE Tracks	Traditional Calendar	A	-2.24	-3.99	-1.16	
			B	*-2.04*	*-2.28*	-0.90	Moderate
			C	-1.27	-2.78	-1.29	
			D	-0.29	-1.48	2.04	
Unstandardized Regression Coefficients for Continuous Classroom Variables				Reading	Math	Language	
13	Proportion from Low Income Home (NSLP) in Class			-0.38	*7.13*	8.15	*Large*
14	Proportion Girls in Class			-3.53	-5.54	-4.02	
15	Proportion AfroAmericans in Class			0.09	-2.34	-4.78	
16	Proportion Asians in Class			-31.14	-17.26	-29.10	Very Large
17	Proportion Hispanics in Class			*-9.02*	-16.69	-21.09	
18	Proportion Other Ethnicity in Class			-64.26	-103.85	-66.18	
19	Proportion Spanish Home Language in Class			11.47	10.74	*60.14*	Very Large
20	Proportion Other Home Language in Class			11.89	11.92	49.98	
21	Proportion New to District in Class			6.50	*9.31*	8.26	*Large*
22	Proportion FEP in Class			-8.96	-13.65	-57.34	Very Large
23	Proportion LEP in Class			-11.18	-0.92	-51.45	
24	Proportion RSP in Class			*8.26*	23.34	14.96	
25	Proportion DIS in Class			-4.85	18.24	11.21	Very Large
26	Proportion GATE in Class			*-10.92*	*-15.47*	*-12.02*	
27	Proportion Overage in Class			*14.90*	*22.51*	*15.51*	Very Large
Cumulative Pct of Total Variance Explained without considering Class Size Reduction				38.0%	34.7%	34.7%	
On Classroom Environment Residual		Reference Category	Level	Reading	Math	Language	
*	Class Size Reduction	No CSR Experience	1996-97 Only	-0.3	0.1	0.5	Negligible
			1997-98 Only	1.0	1.0	0.9	Negligible
			2 Years CSR	*0.8*	0.7	0.7	Negligible
Pct of Total Residual Variance Explained by Class Size Reduction				0.6%	1.0%	0.3%	

student's own special education classification, moving from zero to 100 percent of any special education category in a single class can move the achievement of that class by up to 25 NCE points (more than two years of academic attainment). Similarly, very high concentrations of overage students in a classroom can lower the class attainment by 15 to 24 NCE points.

Large impacts on class performance also result from the proportion of poor students in the class, the proportion of girls in the class and the proportion of students with non-English home languages. A concentration of girls in the class adds to the already significant advantage girls have in test performance. Surprisingly, concentrations of LEP students in class appears to have the effect of offsetting the negative individual level impact of this variable (except in mathematics where the concentration adds further to the LEP students' disadvantage). Where classrooms have high concentrations of poor students, there are large negative consequences for performance in math and language, but no apparent effect (beyond an individual student's own poverty status) on reading scores.

Attendance on various year-round education tracks and being in combination grade classes have Small to Moderate impacts on measured attainment. All year-round tracks suffer in comparison with students attending traditional calendar schools; least so, however, with Track D. Reliance

on combination grade classes for instruction appears to have a small positive impact on student learning in reading and language.

A fourth review of the possible impacts of CSR implementation on student achievement, following the removal of biases created by the classroom context variables, is shown near the bottom of Table 5d. Once again, the pattern of effects changes substantially, revealing that what previously appeared to be class size effects were actually produced because of differences in the composition and structure of the large and small classes. The apparent negative impact on students experiencing small class sizes only during the first year of implementation largely disappears (there is still a trivial -.32 NCE point lower performance reading, offset by a similarly small .49 point positive differential in language, with math scores being virtually identical to the large class math scores). Students participating in reduced size classes only in the second year of implementation appear to have some what better performance when the classroom context variables are removed, but these classes show only about a 1 NCE point advantage over the students with no reduced size class experience. This much achievement gain is about what would be expected from a week or two of additional instructional time.

Classroom environment interactions. As with the analysis of student demographic and classroom assignment variables, we found statistically powerful interactions among the various classroom context variables. Year Round Education track achievements, for example, are substantially connected to the concentration of various ethnic groups in the YRE classes. There were no obvious patterns across SAT-9 subtests, specific YRE tracks, or particular ethnic groups, however. Thus, we have treated this interaction (which accounts for about 1.25 to 2.5 times as much student achievement variance as does CSR itself) as merely random "noise" which must be statistically controlled in order to see the true effect of CSR. Several other interactions at the classroom context level are statistically reliable, but only special education by track assignment has an explanatory power rivaling or exceeding that of CSR.

How Teacher Characteristics Influence Achievement

As shown in Table 5e, one teacher characteristic variable, contract status, has a Very Large impact on achievement. Two other teacher variables, education and ethnicity, have Moderate to Large effects, while teacher gender has a Small impact on student achievement. Teacher contract status has the largest impact, with all teacher groups suffering by comparison with the fully tenured teachers. Temporary contract teachers have the lowest achieving students – up to half a standard deviation (in mathematics) below the levels reached by tenured teachers. We are quick to acknowledge, however, that this does not necessarily mean that the temporary teachers are less effective. They are, after all, likely to get less desirable assignments, and to be bumped from preferable classroom assignments by more senior teachers.

Table 5e also shows that teachers with more education have classes with higher achievement than those lacking training beyond the bachelor's degree. Teachers with advanced degrees have classes where students have test scores that are up to 4 NCE points higher than the BA only teachers. This is about two-tenths of a standard deviation, about the same size as the effect attributed to class size reduction found by researchers studying Tennessee's Project STAR. This

Table 5e. Differences in Mean Achievement Related to Demographic and Schooling factors for 26,126 Students in Eight School Districts, in Grades 2 through 4

Numbers are mean differences for Stanford Achievement Test, NCE Total Reading, Mathematics, Language Scores
(Cell entries are Factor Level NCE's minus the Reference Category identified for each Factor, Unstandardized regression coefficients for continuous variables)
Statistically Significant Values are: **Bold** $p < .001$, ***Italic*** $p < .01$, *Italic* $p < .05$

			Grand Mean:-	42.80	44.78	44.96	
			Std. Deviation:-	21.07	22.01	21.47	Overall
Teacher Characteristic Variables		Reference Category	Level	Reading	Math	Language	Effect Size
28	Contract Status	Tenured	Probationary	-1.64	-2.12	-1.80	
			Temporary	-9.54	***-12.49***	-8.77	Very Large
			Other Contract	-1.00	*-3.54*	-0.96	
29	Education Level	BA/BS	BA+30	1.15	*3.01*	0.73	Moderate
			MA or Greater	*4.04*	4.79	1.78	
30	Teacher Ethnicity	White	Black	***2.36***	***3.42***	***2.06***	
			Hispanic	5.44	4.09	6.06	*Large*
			Other	-0.97	-0.13	0.71	
31	Teacher Gender	Female	Males	*-2.14*	-1.36	-3.24	Small
32	Experience	(NCE points/year)		0.03	0.05	0.00	Negligible
33	Age	(NCE points/year)		0.02	0.03	0.01	
Cumulative Pct of Total Variance Explained without considering Class Size Reduction				38.6%	36.4%	35.2%	
On Teacher Characteristics Residual		Reference Category	Level	Reading	Math	Language	
*	Class Size Reduction	No CSR Experience	1996-97 Only	-0.5	-0.1	0.3	Negligible
			1997-98 Only	1.0	1.0	0.8	Negligible
			2 Years CSR	*0.8*	0.7	0.7	Negligible
Pct of Total Residual Variance Explained by Class Size Reduction				0.1%	0.1%	0.0%	
Cumulative Pct of Total Variance when School and District Levels are Included:-				39.6%	37.4%	36.6%	

finding that less well educated teachers have students with significantly lower test scores tends to reinforce the common sense notion that teachers with advanced degrees have more advanced professional skills and abilities. It is possible, however, that this effect is caused by the fact that better educated teachers are more attractive employees and thus have the capacity to select classes with easier to teach or higher performing students.

The Moderate relationship between teacher ethnicity and student achievement offers some interesting clues as to how teachers might impact student learning. After controlling for other teacher characteristics, the demographic characteristics of their students, and classroom assignment and context variables, we find that student achievement tends to be higher in the classrooms of *non-White* teachers. Both African American (Black) and Hispanic teachers have classes that do better than would otherwise be predicted, when compared to their White counterparts.

Female teachers appear to have a Small, but significantly positive impact on the production of reading and language test scores. Men and women teachers have about the same level of impact on mathematics achievement.

Neither teacher age nor experience make any significant contribution to student test scores, once the effects of other factors have been eliminated.

CSR impact after controlling for teacher characteristics

Shown near the bottom of Table 5e, is our final analysis of CSR impacts on student achievement. Removal of the biasing effects resulting from the fact that small and large classes were being served by teachers with different characteristics does not change the picture very much. Students

BEST COPY AVAILABLE

with small class experience only during the first year of CSR implementation have trivially lower achievement in reading and mathematics than the students with no small class experience, but the other two CSR treatment groups display slightly higher scores than the no CSR experience group on all three SAT-9 subtests. In no case, however, does the difference between students in small classes and those in large classes differ by more than 1.2 NCE points. The average gain for the second-year implementation group (the group with deriving the greatest benefit from CSR) is about 1.01 NCE points – less than 5 one-hundredths of a standard deviation, about the amount of achievement gain expected from two weeks of student maturation and school instruction.

Teacher characteristic interaction effects. Table 9 and Figure VII show how the interaction between teacher ethnicity and contract status are related to student achievement in reading. As this table and figure reveals, the biggest differences in teacher influence over student achievement are between tenured White teachers and White teachers with other types of contracts. Tenured African American (Black) teachers and Hispanic teachers are close in their performance to Whites. And probationary contract African American teachers are very close in their contributions to student achievement to tenured White teachers. Among "temporary" and "other" contract holders, there are too few teachers in the non-White categories to compare, but there is a decided loss in student achievement associated with these irregular contract status among White teachers.

Figure VII. Interaction of Teacher's Contract Status with Ethnicity on SAT-9 Total Reading Achievement

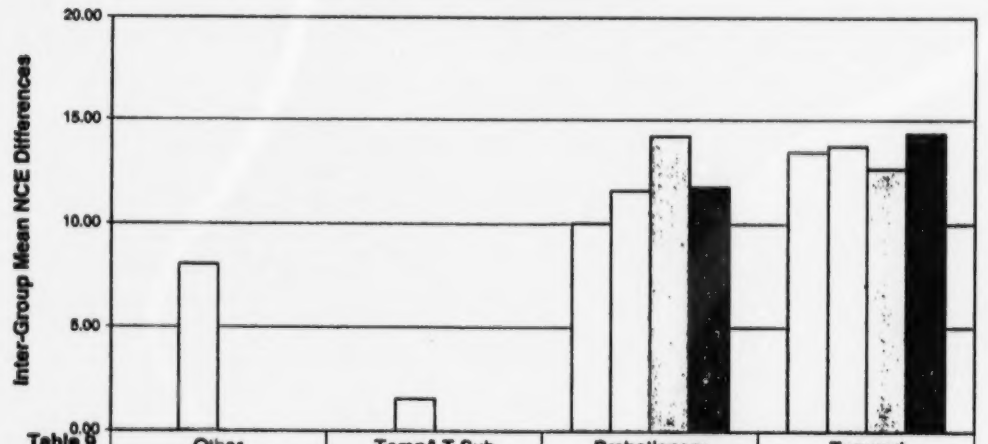


Table 9

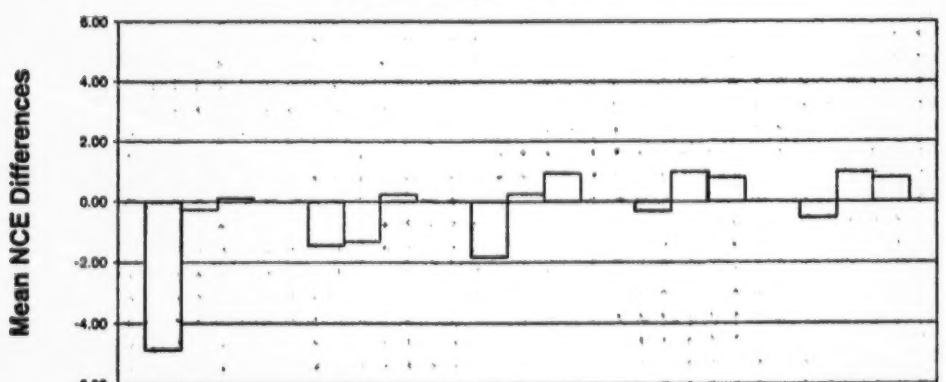
	Other	Temp/LT Sub	Probationary	Tenured
Other			10.01	13.43
White	8.03	1.57	11.58	13.73
Black			14.21	12.64
Hispanic			11.73	14.32

Summarizing the Impact of CSR on Overall Student Achievement

Figures VIII through X summarize the CSR impact analysis discussed above. The graphs shown in these figures depict the extent to which the three CSR class types (small classes in 1996-97 only, in 1997-98 only and in both years) yielded SAT-9 test scores above (or below) those generated by students with no reduced class size experience. In each figure, the first set of bars represent the amount of difference in performance found in the original SAT-9 data, without controlling for any of the potentially confounding factors. The second cluster of three bars represents the deviations when the effects of student demographic factors are controlled. The third cluster represents the deviations found when classroom assignment factors are removed; the fourth cluster shows the results upon removal of the effects of classroom context variables. The final cluster of bars represents the deviation scores left when all of the potentially biasing factors have been statistically removed from the data.

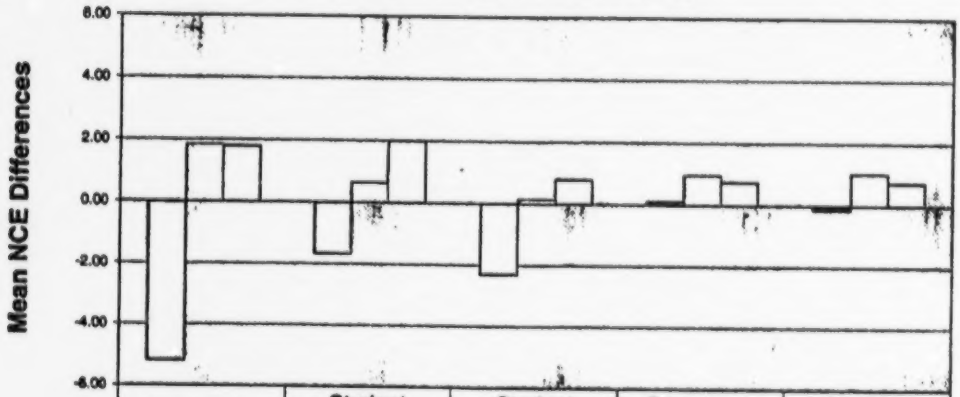
The data on which these graphics are based strongly support two conclusions. First, during the first two years of implementation, California's CSR initiative has probably made a very small, but positive contribution to raising students' achievement. The contribution is so small, however, and so entangled with various demographic, classroom and teacher variables that we cannot be certain it will be reliably maintained in future years. Moreover, virtually all of the benefit found in this study sample accrued to students who participated during the second year of implementation, regardless of whether or not they also participated during the first year. It seems quite possible that first year implementation was so rushed and so disruptive to established school routines that potentially positive effects were dragged down by implementation problems. For whatever reason, students participating only during the first year showed no achievement gains relative to students with no reduced class size experience.

Figure VIII. SAT-9 Total Reading by CSR Experience for All Students in Eight District Study



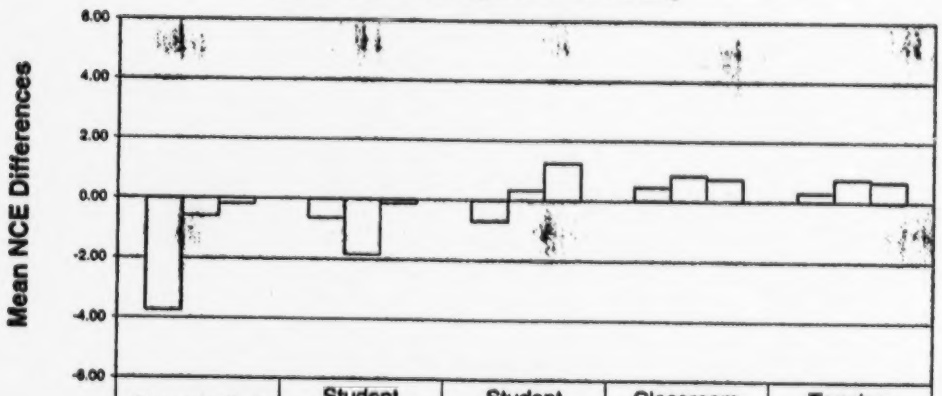
	Uncontrolled	Student Demographics	Student Assignments	Classroom Context	Teacher Characteristics
□ 97 Only	-4.86	-1.45	-1.83	-0.32	-0.54
□ 98 Only	-0.27	-1.32	0.23	0.99	0.99
□ 97 & 98	0.11	0.22	0.94	0.79	0.79

Figure IX. SAT-9 Total Mathematics by CSR Experience for All Students in Eight District Study



	Uncontrolled	Student Demographics	Student Assignments	Classroom Context	Teacher Characteristics
☐ 97 Only	-5.18	-1.67	-2.30	0.11	-0.12
☐ 98 Only	1.84	0.64	0.15	0.97	1.02
☐ 97 & 98	1.79	1.99	0.79	0.74	0.74

Figure X. SAT-9 Total Language by CSR Experience for All Students in Eight District Study



	Uncontrolled	Student Demographics	Student Assignments	Classroom Context	Teacher Characteristics
☐ 97 Only	-3.75	-0.65	-0.71	0.47	0.31
☐ 98 Only	-0.60	-1.85	0.33	0.85	0.76
☐ 97 & 98	-0.19	-0.13	1.24	0.73	0.68

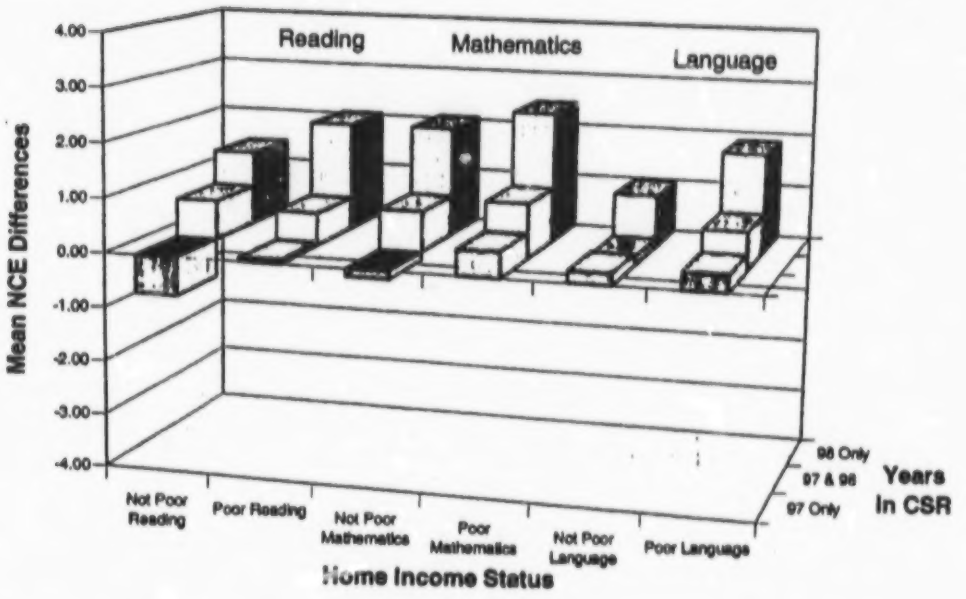
Second, after the effects of other variables are controlled, the CSR effect is not more than one NCE point, which is no more than would be expected from about two weeks of student maturation and school instruction. The story is essentially the same for each of the three SAT-9 subtests examined – virtually all of the explainable variations in student achievement are due to factors other than CSR implementation.

Step 3: Exploring the Possibility that CSR has Benefits for Specific Student Groups

Having examined the small but positive impact of CSR on the average academic performance of all students, we turn now to examination of whether smaller class sizes have been beneficial to specific groups of students. To do so, we divide our sample along three of the most important factors differentiating students into important subgroups – poverty, ethnicity and academic ability – and test whether CSR experience affects the resulting subgroups differently. Since one of our eight districts does not designate GATE children and has very few poverty students, this part of our analysis was undertaken using data from only seven districts (the total student sample dropped from 26,126 to 24,176).

Differential CSR impact on poor and not poor students. The first test of differential impact was performed by dividing the sample between the poor and the not poor students (i.e., separating those on the National School Lunch Program from those who are not). Figure XI graphically depicts the achievement of poor and not poor students on each of the three SAT-9 tests (reading, math and language) after the effects of all demographic, class context and teacher characteristics variables have been statistically controlled. A review of the bars on this graph quickly tells the story: poor students derived very slightly, but consistently, greater benefits from the small class experience in each of the three CSR class types (i.e., 1996-97 only, 1997-98 only, and both years). The benefit for the 1997-98 experience is reliably different from that for students who had no CSR experience. In no case are the differences between the poor and the not poor students statistically significant, so we cannot say with any confidence that these differences would continue to be displayed in future tests. It is possible to be reasonably confident, however, that poor children will benefit at least as much as their not poor peers.

Figure XI. SAT-9 Reading, Mathematics, and Language Achievement Comparing CSR Experience against No CSR Experience Across Poverty Status and Years when CSR Experience Occurred

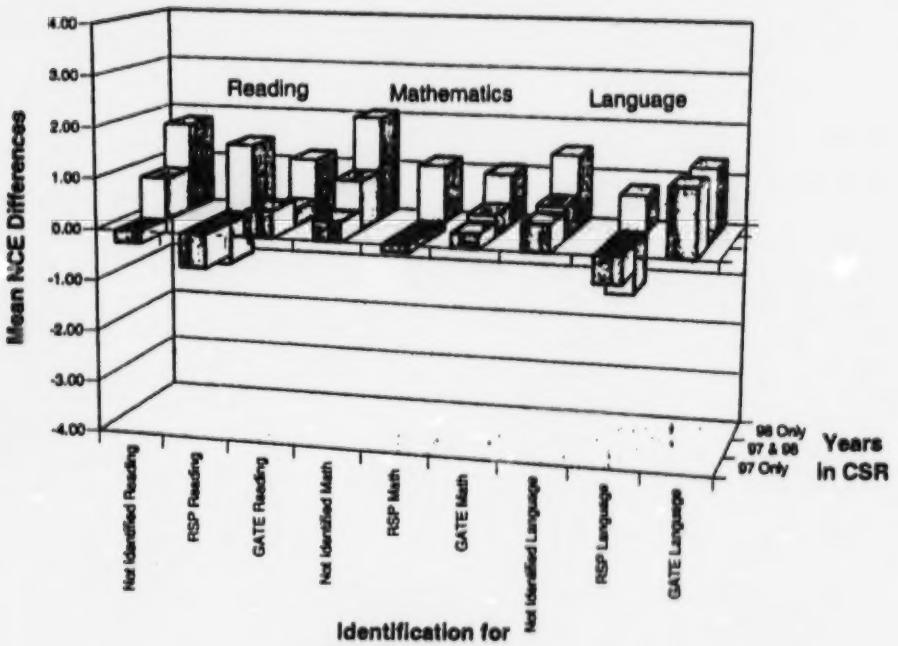


	Not Poor Reading	Poor Reading	Not Poor Mathematics	Poor Mathematics	Not Poor Language	Poor Language
■97 Only	-0.70	0.07	-0.14	0.46	0.21	0.31
□97 & 98	0.72	0.60	0.75	1.01	0.17	0.73
□98 Only	1.37	2.02	2.04	2.41	0.96	1.83

Differential CSR impact on children at different academic performance levels. We next tested whether the lower achieving children assigned to Special Education Resource Specialist (RSP) Programs or the high achieving children certified for Gifted and Talented Education (GATE) programs display different responses to CSR than their normal achievement range classmates. Figure XII graphically represents the extent to which students in each of the three CSR class types (i.e., only in 1996-97, only in 1997-98, and both years) differed from those

classes where student had no CSR experience. The left three rows of bars in the figure present the reading test results. The results for the 1996-97 only group are not consistent with the other two conditions, but for those whose CSR experience included 1997-98 participation, there was a

Figure XII. SAT-9 Reading, Mathematics, and Language Achievement Comparing CSR Experience against No CSR Experience across Special Education Status and Years when CSR Experience Occurred

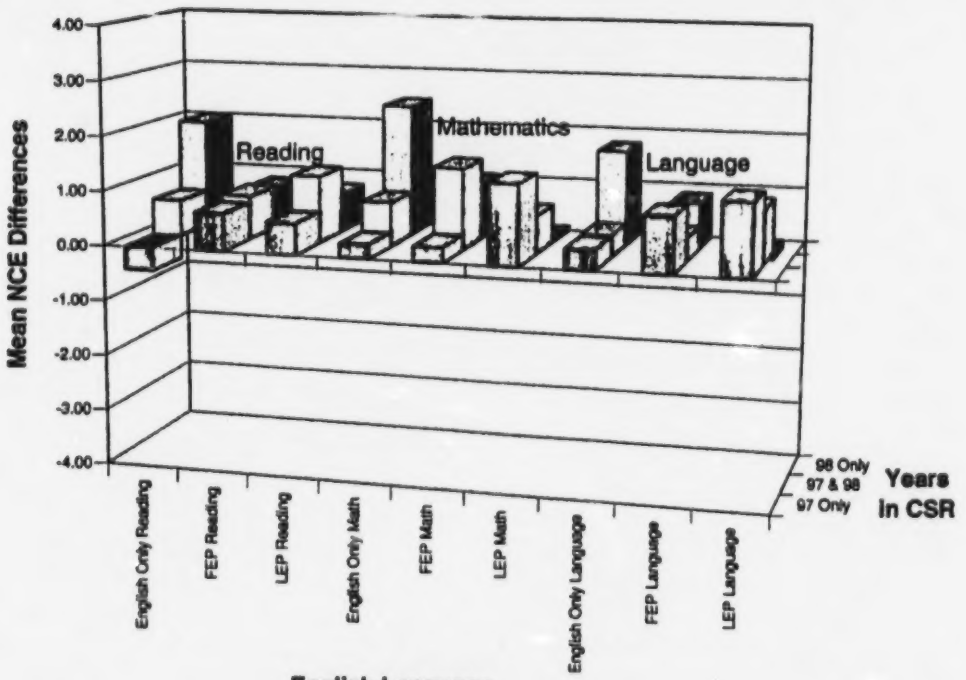


	Not Identified Reading	RSP Reading	GATE Reading	Not Identified Math	RSP Math	GATE Math	Not Identified Language	RSP Language	GATE Language
■97 Only	-0.23	-0.65	0.51	0.37	-0.13	0.28	0.52	-0.53	1.32
□97 & 98	0.85	-0.81	0.33	0.97	-0.15	0.42	0.64	-0.98	0.10
□98 Only	1.76	1.38	1.14	2.04	1.13	1.01	1.44	0.71	1.29

very slight achievement advantage for the middle level students who were neither in GATE nor in RSP programs. In the center of Figure XII are the bars reporting results for the SAT-9 mathematics test. In this case, the results are consistent across all CSR class experiences – students not certified for either GATE or RSP programs had a consistent, but statistically

insignificant, greater benefit from participation in reduced size classes. The performance of not identified students' in 1997-98 is reliably different from not identified students who had no CSR experience. The language story is like that for reading, and is shown toward the right side of the figure. In language, for students whose CSR experience included small class assignments in 1997-98, a slightly larger, but statistically insignificant, benefit went to the middle achieving students. The GATE students showed the most benefit from CSR participation among students participating only in the first year of implementation. Again, we caution that none of the differences depicted on these graphs are statistically significant and we can have no confidence that they will be repeated in further studies of how students of differing academic performance levels benefit from smaller classes.

Figure XIII. SAT-9 Total Reading, Mathematics, and Language Achievement Comparing CSR Experience against No CSR Experience across English Language Proficiency Status and Years when CSR Experience Occurred

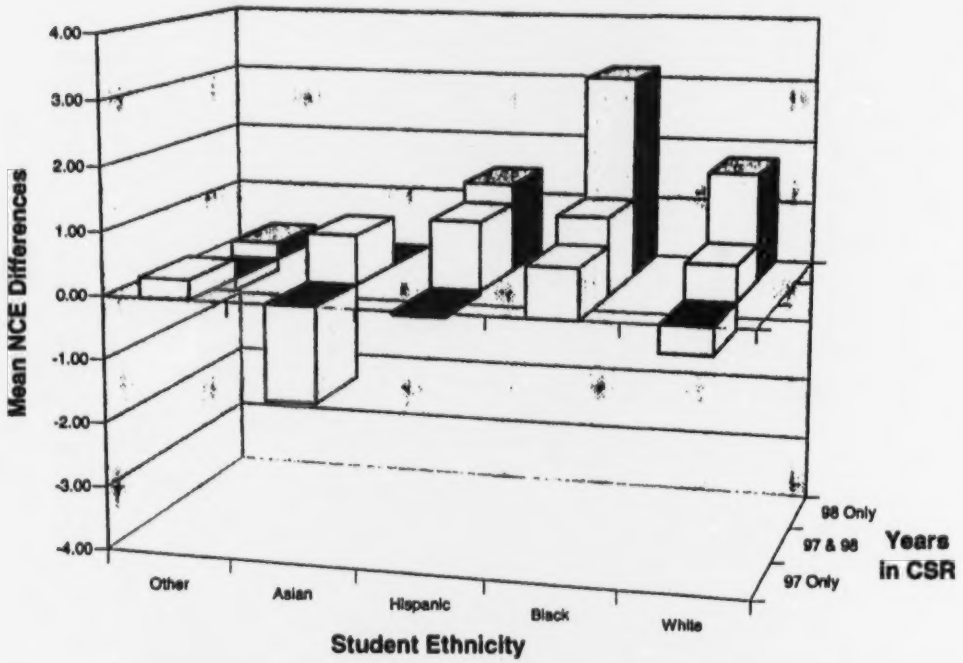


	English Only Reading	FEP Reading	LEP Reading	English Only Math	FEP Math	LEP Math	English Only Language	FEP Language	LEP Language
97 Only	-0.42	0.68	0.57	0.26	0.28	1.47	0.38	1.01	1.32
97 & 98	0.67	0.81	1.23	0.81	1.50	0.70	0.40	0.42	0.99
98 Only	1.97	0.77	0.71	2.42	1.02	0.08	1.73	0.82	-0.12

Differential CSR impacts on students with differing levels of English language proficiency. Figure XIII compares the marginal benefits of CSR participation for the major for students who were English proficient only, designated as fluent English proficient (FEP), and designated limited English proficient (LEP). Similar to Figures XI and XII, the reading, mathematics, and language test group means are arrayed in three groups of three from left to right in the figure. The only consistent message for students of varying English language proficiency is that English only students had higher achievement when their CSR experience was in 1997-98 only. These results are statistically reliable for the English only students, but they are not significantly different from FEP and barely reliably different from LEP students. Unlike for home income status and special education status, there is no strong pattern of successive benefit across the CSR implementation categories. But since these small values are not statistically reliably different from no CSR experience, it is advisable to not attempt to say more than when there is a benefit, it appears to be going to students who had their CSR experience only in 1997-98, and like those for students not identified for special education, the reliability of these results is not so high that generalizations should be made to a population outside of the sample studied. Further, with the inconsistency from year to year, caution should be taken when attempting to predict outcomes for any successive year's experiences for students for whom their home language is not English.

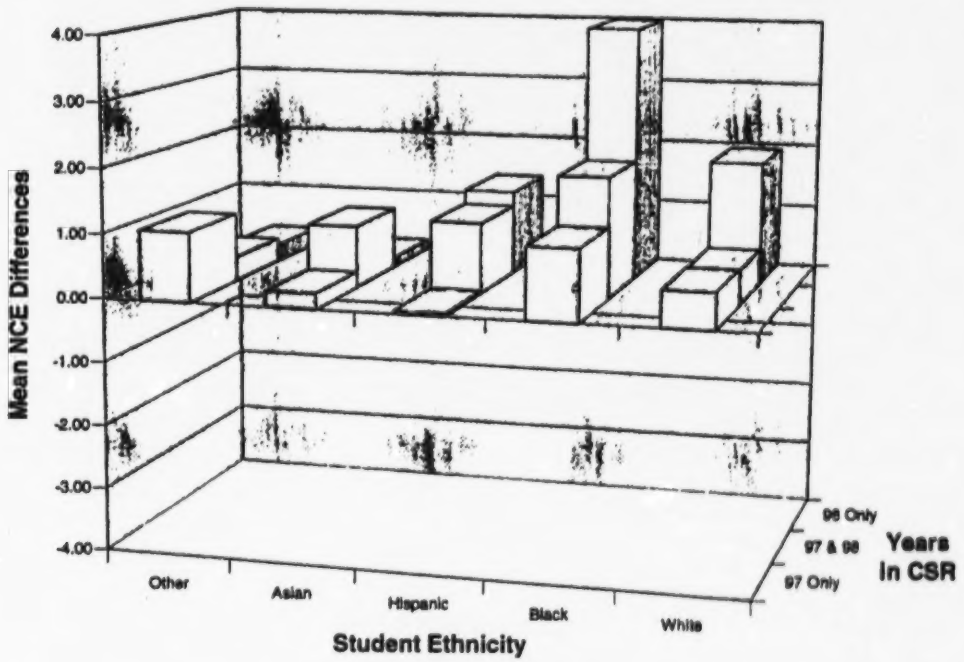
Differential CSR impacts on students from differing ethnic groups. Figures XIVa, XIVb, and XIVc compare the marginal benefits of CSR participation for the major ethnic groups represented in our study sample. Figure XIVa presents the reading results. This figure shows that African American (Black) students showed a consistently higher benefit from CSR experience in each of the three class types (i.e., 1996-97 only, 1997-98 only and both years). The mathematics results show the same consistent pattern – African American students got consistently higher benefits from their reduced size class experiences – especially those who participated only in the second year of implementation (1997-98). In the case of the SAT-9 language subtest, presented in Figure XIVc, African American students consistently benefited more than their White and Hispanic classmates, but Asians and Others benefited a little more in some treatments contexts. Only African American and White students had 1997-98 only CSR experience group means that were reliably different from those students of the same respective ethnicity who had no CSR experience. Importantly, in the reading and mathematics cases, the marginally higher benefits of CSR accruing to African American students is statistically reliable as well, indicating that this result would probably be replicated in further testing. Moreover, in the case of mathematics, the 1997-98 year benefit to African American students participating only in the second year of CSR implementation was a substantial 3.97 NCE points, which represents a quarter to a third of a grade level improvement in performance. Of course, statistical reliability does not necessarily mean that further analysis will see results that are this large, but they will almost certainly continue to be positive. Positive effects for African American students were found by Bingham (1994) and Konstantopoulos (1999) in their re-analyses of the Tennessee data.

Figure XIVa. SAT-9 Total Reading Achievement Compared to No CSR Experience for Interaction of Student Ethnicity with Years CSR Experience Controlled at the Student, Classroom & Teacher Levels



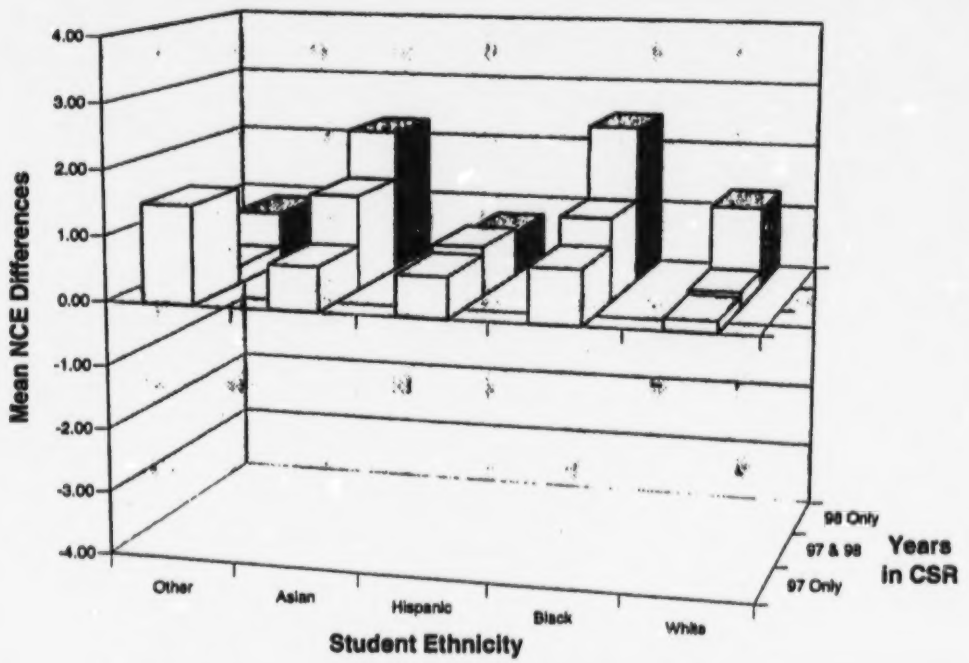
	Other	Asian	Hispanic	Black	White
☐97 Only	0.32	-1.49	-0.06	0.79	-0.42
☐97 & 98	-0.19	0.79	1.09	1.23	0.58
☐98 Only	0.27	-0.01	1.41	3.16	1.71

Figure XIVb. SAT-9 Total Mathematics Achievement Compared to No CSR Experience for Interaction of Student Ethnicity with Years CSR Experience Controlled at the Student, Classroom & Teacher Levels



	Other	Asian	Hispanic	Black	White
☐97 Only	1.08	0.24	0.05	1.12	0.56
☐97 & 98	0.38	0.97	1.11	1.89	0.56
☐98 Only	0.28	0.16	1.32	3.97	1.92

Figure XIVc. SAT-9 Total Language Achievement Compared to No CSR Experience for Interaction of Student Ethnicity with Years CSR Experience Controlled at the Student, Classroom & Teacher Levels



	Other	Asian	Hispanic	Black	White
□97 Only	1.51	0.69	0.63	0.84	0.18
□97 & 98	0.27	1.46	0.76	1.28	0.25
□98 Only	0.82	2.23	0.69	2.43	1.24

Summary, Conclusions and Recommendations

This report presents a comprehensive preliminary analysis of how California's Class Size Reduction (CSR) initiative has impacted student achievement during the first two years of implementation. The analysis is based on complete student, classroom and teacher records from 26,126 students in 1,174 classrooms from 83 schools in 8 southern California school districts. The data include reading, mathematics and language test scores from the Stanford Achievement Test (Version 9 – SAT-9) collected through California's Star testing program. Also analyzed are 34 variables covering student demographics, school assignments, classroom contexts, and teacher characteristics. The evidence reviewed supports nine broad conclusions and leads to five recommendations to education professionals and policy makers.

> Conclusion #1: CSR is massive, expensive and adopted in conjunction with a complex array of other new policy initiatives aimed at improving California school performance. Evaluating the impact of this initiative is made particularly difficult by the fact that so many other important initiatives are being simultaneously pursued.

At a direct cost exceeding $2.3 billion in the first two years of implementation, CSR is the most expensive reform of public education ever undertaken in California (California Department of Education 1999). There are many reasons for believing that CSR may be helpful to public education. Improving student achievement is certainly its most important goal, however. Thus, student achievement effects of CSR implementation are the focus of this report. CSR was not adopted as an experiment or as a test of how much it could contribute to student performance, but was implemented comprehensively and on very short notice. Moreover, CSR was adopted at the same time as revisions in teacher preparation, mandates for reforming bilingual education, development of new curriculum frameworks and materials, adoption of a new statewide test, development of a new performance accountability system and numerous other policies whose effects cannot be precisely estimated. It may never be possible to know with certainty how much this initiative has contributed to student learning.

> Conclusion #2: Rapid implementation of California's CSR initiative placed substantial stresses on school facilities, created an intense demand for new teachers, and encouraged a shift to Year Round school calendars to accommodate enrollment growth and reduced size classes.

These stresses are quite likely to mean that CSR is functioning differently during its first few years of operation than can be expected in the years ahead. Schools hired many more teachers who are not fully credentialed and who lack training comparable to the average teacher in the years immediately prior to CSR implementation. An earlier study by the California Educational Research Cooperative documented a sharply elevated frequency of first-year and not fully qualified teachers serving in reduced size classes (Ogawa and Stine 1998).

> Conclusion #3: School officials were faced with tough decisions regarding the sequence of CSR implementation and the allocation of opportunities to participate in reduced size classes on the part of teachers and students.

As a consequence of the choices made, students and teachers were definitely not randomly distributed among large and small classes. Of 34 variables examined in this study, only student gender did not significantly relate to whether students were assigned to large or small classes for one or both of the first two years of CSR implementation. Since academic achievement is influenced by multiple layers of demographic influences, classroom assignment variables, school and classroom contexts, and teacher characteristics, any effort to evaluate the impact of CSR must carefully attend to the imbalances in student and teacher participation.

Conclusion #4: Implementation biases responsible for differences in student and teacher participation in reduced size classes were strikingly different in the first and second years of CSR implementation.

Students in reduced size classes during the first year were more likely to be from ethnic minority groups, from poor neighborhoods and attending year round schools than those first participating in the second year of implementation. Students who did not have access to reduced size classes until the second year were more likely to be new to the district in 1998, and to come from English speaking homes.

Conclusion #5: Statistical analyses revealed that biases in CSR participation are sufficiently strong that knowing the demographic, school assignment and teacher characteristics of any given student makes it possible to substantially predict whether they were in small or large classes for one or both years.

Specifically, multiple discriminant analysis of CSR implementation biases improves by more than 35 percent our ability to predict their CSR experience. This means, quite simply, that achievement differences between the large and small classes created by California's CSR may be, to a substantial degree, determined by differences in who has participated, rather than how class size itself affects learning.

Conclusion #6: The factors associated with the biases in student participation in various CSR implementation alternatives are, themselves, much more strongly related to student achievement than is class size reduction.

Twenty-five of the 34 variables examined in this study were at least as powerful as CSR experience in predicting student achievement. Of these variables, student poverty, gender, ethnicity, home language, special education certification and transiency are two to twenty times as powerful as CSR experience in predicting student achievement. Additionally, teacher contract status, ethnicity, education level and gender are from two to ten times as powerful as CSR experience in predicting student achievement. As a result, relatively small biases in the assignment of students or teachers to small classes can create outcome differences that are as large or larger than the CSR effect.

Conclusion #7: Nevertheless, after controlling for all of the available biasing factors, there remains a small positive impact from CSR on student achievement as measured by the Stanford-9 achievement test. The CSR impact varies from year

to year, however, and is not consistent across the reading, mathematics and language subtests of the SAT-9.

After statistically removing the effects of the known biasing variables, CSR experience during the first two years of implementation accounted for about a 1 NCE point gain on the 1998 SAT-9 tests. This amount of achievement gain is approximately as much as would be expected to result from about two weeks of additional student maturation and instruction. The CSR contribution accounts for about one-tenth of one percent of all student achievement variation, whereas the other variables analyzed in this report account for 35.2 to 38.6 percent of student achievement variance.

Conclusion #8: The benefits of CSR experience are apparently not evenly distributed among student groups. African American (Black) students showed stronger gains in achievement associated with small class experience than did other ethnic groups. There is weaker evidence that poor students and children not certified for special education may benefit slightly more from participation in reduced size classes than to those who are not poor or are certified for special education.

Again, these findings represent the marginal contributions of CSR, after controlling for the other factors that influence student achievement. Only in the case of the African American students do inter-group differences reach the level of statistical reliability needed to be confident that the differences found in this study sample would be confirmed in further tests.

Conclusion #9: Because class size reduction is so deeply entangled with student, school and teacher variables, it virtually impossible to fully disentangle the various factors influencing achievement with the usual *post hoc* exploratory data analysis.

Adequate assessment of the influence of CSR on student achievement will require a convincing conceptual framework capable of directing attention to the specific mechanisms by which CSR is expected to raise student performance. Absent a compelling theory of the mechanisms of performance improvement, it is impossible to know with any degree of certainty which of the very powerful factors examined in this report need to be controlled through planned variation, randomized implementation, or statistical methods when interpreting the data.

Recommendations for Policy Action

The five policy recommendations representing logical extensions of the data analyses presented in this report include:

Recommendation #1: The most obvious implication of this study is that California would be well advised to "stay the course" with class size until its full effect can be analyzed and documented.

Initial implementation has almost certainly been sufficiently disruptive of school operations that the data analyzed here do not tell the whole story of what can be expected from class size reduction. Until we are able to see how much the academic performance of California's fourth

graders can be improved by up to four years of smaller class size experience, it is not appropriate to assert that we really know that CSR does or does not improve student achievement. As Murnane and Levy (1996) found in Austin, Texas, highly effective schools required four years to see consistent growth in achievement from their simultaneous introduction of class size reduction with other major instructional programs.

> Recommendation #2: Given the evidence of rather limited impact during the first two years of implementation of CSR, it is appropriate to begin now testing whether substantial investments in targeted student intervention programs, or expanded professional development activities might contribute more to student learning than a simple reduction in the number of children assigned to a classroom.

Since school program assignments, year round school track assignments, segregation of student groups within the schools, and teacher education and contract status are all more powerfully correlated with student achievement than CSR, it would seem reasonable that policies and programs be developed on the basis of careful examination of how these factors are influencing student learning and how they might be managed to better capitalize on their benefits.

> Recommendation #3: Support needs to be given to work that establishes appropriate explanatory frameworks for interpreting the relationship between class size and student achievement.

To date, research on the relationship between class size and student achievement has been remarkably devoid of meaningful theory. Exactly why removing some children from a classroom should cause the achievement of those remaining to go up remains largely unexplained, even as it is widely expected to be more true than careful data analysis has been able to support.

From the nature of the policy debates informing the adoption of CSR, and from the approaches taken in most research studies, we can infer that there are four competing theoretical frameworks for explaining how smaller classes might be expected to improve school performance. The first, and most common, framework assumes that CSR will work because it increases the instructional resources available to each child in the school. It is assumed that lowering the number of children in a classroom will mean that each child has more access to the teacher and probably more physical space. As educators or policy makers realize that CSR may have less impact than initially hoped, they start to focus on whether teachers need to change their instructional practices in order to produce the benefits expected from smaller classes. That is, they begin to hypothesize that additional resources alone will not produce results – changed instructional practices, possible only in smaller classes, are required. This instructional change model sees CSR as an opportunity to improve schooling, but one that will only be realized if teachers adopt instructional practices appropriate to the smaller class context. The research literature is not very clear about exactly what instructional changes are needed, and even less clear about why some teachers are more likely to make the appropriate changes than are others.

A third theoretical framework sees CSR as changing classroom organization rather than resources or instructional techniques. This view hypothesizes that smaller classes raise

achievement by creating more homogeneous classroom groups and by reducing the frequency with which teachers have to cope with students' learning problems. The fourth theoretical model extends the idea of CSR impact on classroom organization by proposing that smaller classes become effective through the creation of greater student engagement and motivation. The working hypothesis behind this fourth view is that the effectiveness of the smaller classes springs from their ability to reduce alienation and enhance the development of a cohesive community among students and teachers. From this point of view, smaller classes are expected to be most effective in improving the learning of those students most often disengaged from the learning process. Thus, children who have educational handicaps, who are stressed by poverty, or who have been the victims of racial or ethnic prejudice are most expected to benefit more than those from mainstream, middle class families.

Each of these theoretical models is a reasonable account of why we should expect class size changes to produce changes in student achievement. No doubt, there are other reasonable theories. It is important to develop these theories to the point that their implications for achievement patterns and interactions with the student, classroom and school level variables reviewed in this report can be conceptualized and tested.

> Recommendation #4: The educational policy community needs to continue the search for school reform and improvement policies that promise to have achievement effects as large as poverty, home language and student ethnicity.

Quite obviously, class size reduction is not the "silver bullet" needed to offset serious educational challenges facing children from poor, minority or non-English speaking homes. Even the most optimistic projections of the achievement gains to be generated through continued and careful implementation of CSR do not lead us to seriously believe that this policy will solve the pressing problems of low achievement haunting California schools.

> Recommendation #5: A serious effort needs to be made to strengthen the ability of education researchers and school professionals to develop data systems capable of supporting analysis of relationships between the implementation of specific educational programs and services and resulting changes in student achievement.

Researchers and school professionals interested in documenting the impact of various programs and policies on student achievement find themselves faced with a continuing and serious problem of data availability and usability. Current educational data systems (such as California's CBEDS and STAR data systems) lack two characteristics that are absolutely essential for documentation of policy effectiveness. First, these data systems typically collect only one of the three elements of a program or policy evaluation. To evaluate any program or policy, basic data on school inputs related to student, classroom and school composition must be linked to information on the actual delivery of educational programs and services. These data must, in turn, be linked to measures of student attainment (or other targeted educational goals). California's CBEDS system provides useful data about student characteristics and the teaching resources made available to them (though the system does not enable anyone to know with any degree of confidence which students had access to what teacher or school resources). The STAR data system provides important, though somewhat limited, data on how well students are

achieving academic outcomes. There is no comparable data system recording what instructional programs or practices were used by schools or classroom teachers in their efforts to educate the students, however. Even more problematic is the fact that data on student achievement and the records of resources used in their instruction are stored in ways that do not permit continued monitoring of the success of failure of specific educational programs and services. Typical data collections maintain records for a year at a time without permitting tracking student performance from year to year, or continuing analysis of resource availability or program and service delivery processes.

California's *Class Size Reduction Program* is certainly in its formative stages, as this evaluation makes clear. Determining the extent to which CSR is creating impacts on student achievement has been and will continue to be a profound challenge. Evidently, the disruptive effects of statewide movement toward class size reduction are preventing schools and districts from stabilizing the educational environment for children and their teachers in reduced size classes. With the current state-of-the-art, a system for linking classroom practices and instructional or curricular influences to separate data systems that track achievement and other individual student, classroom, teacher, and school factors is not to be found. The circumstances defining where student achievement occurs can not currently be connected to information about the processes that produce student achievement (at least not in any cost-effective manner). As such, it remains a matter of theoretical investigation to determine what mechanisms operating in or as a result of reduced size classes might lead to particular outcomes. Further efforts to evaluate the impact of CSR or any other policy affecting classroom composition and activity will require the simultaneous development of more sophisticated data systems and clearer conceptions of how policies ought to lead to patterns of student achievement or any other targeted outcome of interest. Fuller knowledge of how teachers and students come to produce higher achievement may very well indicate that smaller class sizes are a critical ingredient. Within the turbulent conditions of the first two years of CSR implementation, there is no evidence that reduced size classes are the key or among the most powerful ingredients for substantially raising student achievement in the short-term.

REFERENCES

Alexander, K. L., Entwisle, D. R., and Thompson, M. S. (1987). School performance, status relations, and the structure of sentiment: Bringing the teacher back in. *American Sociological Review, 52*(5), 665-682.

Balow, I. H., and Schwager, M. (1990). *Retention in grade: A failed procedure (AY-001).* Riverside, CA: California Education Research Cooperative, School of Education, University of California, Riverside.

Bingham, C. S. (1994). *Class size as an early intervention strategy in white-minority achievement gap reduction.* Paper presented at the Conference-within-a-Conference, American Association of School Administrators, San Francisco, California. (ERIC Document No. ED 368077)

Bohrnstedt, G. W., and Parrish, T. B. (1998). *California's class size reduction initiative: Is it likely to reduce or create further inequities in California districts?* Paper presented at the Annual Meeting of the American Sociological Association, San Francisco, California.

Burns, R. (1996). *A study of combination class achievement (SA-006).* Riverside, CA: California Education Research Cooperative, School of Education, University of California, Riverside.

Burns, R. B., and Mason, D. A. (1998). Class formation and composition in elementary schools. *American Educational Research Journal, 35*(4), 739-772.

California Department of Education (1999). *Class size reduction: Finger tip facts about CSR.* Sacramento, CA: California Department of Education. (http://www.cde.ca.gov/ftpbranch/sfpdiv/classize/facts.htm)

California Department of Education (1996). *Class size reduction packet 1996-97.* Sacramento, CA: California Department of Education.

Egelson, P, Harman, P., and Achilles, C. M. (1996). *Does class size make a difference? Recent findings from state and district initiatives.* Greensboro, NC: Southeastern Regional Vision for Education (SERVE). (ERIC Document No. ED 398644)

Entwisle, D. R., Alexander, K. L., and Olson, L. S. (1997). *Children, schools, and inequality.* Boulder, CO: Westview Press.

Finn, J. D., and Achilles, C. M. (1998). *Tennessee's class-size study: Questions answered, questions posed.* Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, California.

Finn, J. D., and Achilles, C. M. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal, 27*(3), 557-577.

Finn, J. D., Fulton, D., Zaharias, J., and Nye, B. A. (1989). Carry-over effects of small classes. *Peabody Journal of Education, 67*(1), 75-84.

Glass, G. V., Cahen, L. S., Smith, M. L., and Filby, N. N. (1982). *School class size: Research and policy.* Beverly Hills, CA: Sage Publications.

Glass, G. V., and Smith, M. L. (1979). Meta-analysis of research on the relationship of class-size and achievement. *Educational Evaluation and Policy Analysis, 1*(1), 2-16.

Glass, G. V., and Smith, M. L. (1978). *Meta-analysis of research on the relationship of class-size and achievement.* Boulder, CO: Laboratory for Educational Research, University of Colorado.

Hedges, L. V. (1999). *The effects of small classes: The Tennessee class size experiment.* Paper presented at the Annual Meeting of the American Educational Research Association, Montréal, Québec, Canada.

Hymon, S. (1997). A lesson in classroom size reduction. *School Planning and Management, 36*(7), 18-23.

Illig, D. C. (1997). *Early implementation of the class size reduction initiative.* Sacramento, CA: California Research Bureau, California State Library.

Konstantopoulos, S. (1999). *Do small classes really help minorities and the disadvantaged?* Paper presented at the Annual Meeting of the American Educational Research Association, Montréal, Québec, Canada.

Krueger, A. B. (1998). *Experimental estimates of education production functions (Working Paper #379).* Princeton, NJ: Princeton University, Industrial Relations Section.

McKeon, R. C. (1992). *Estimating cost functions and measuring wealth neutrality of 1990-91 class size reduction implementation in selected Oklahoma school districts.* Unpublished doctoral dissertation, The University of Tulsa, Tulsa, Oklahoma.

Maier, P., Molnar, A., Percy, S., Smith, P., and Zahorik, J. (1997). *First year results of the student achievement guarantee in education program.* Milwaukee, WI: University of Wisconsin, Milwaukee, Center for Urban Initiatives and Research.

McRobbie, J. (1998). Fresno parents win fight against year-round schooling. *WestEd policy support program: Class size update page.* San Franciso, CA: WestEd (http://www.wested.org/policy/class_size/feb-march98_update.htm).

Mitchell, D. E., Beach, S. A., and Badarak, G. (1989). Modeling the relationship between achievement and class size: A re-analysis of the Tennessee Project STAR data. *Peabody Journal of Education, 67*(1), 34-74.

Mitchell, D. E., Carson, C., and Badarak, G. (1989). *How changing class size affects classrooms and students (CS-002).* Riverside, CA: University of California, Riverside, California Educational Research Cooperative.

Mitchell, D. E., Destino, T., and Karam, R. (1997). *Evaluation of English language development programs in the Santa Ana Unified School District: A report on data system reliability and statistical modeling of program impacts (SA-008).* Riverside, CA: University of California, Riverside, California Educational Research Cooperative.

Mitchell, R. E., and Mitchell, D. E. (1999). *Organizational segregation of student achievement within elementary schools: The influence of multi-track year-round schools.* Unpublished manuscript, California Educational Research Cooperative, School of Education, University of California, Riverside, California.

Molnar, A. (1998). *Smaller classes not vouchers increase student achievement.* Harrisburg, PA: Keystone Research Center.

Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children, 5*(2), 113-127.

Murnane, R. J., and Levy, F. (1996). Evidence from fifteen schools in Austin, Texas. In G. Burtless (Ed.), *Does money matter? The effect of school resources on student achievement and adult success* (pp. 93-97). Washington, DC: Brookings Institution Press.

Ogawa, R. T., and Stine, D. (1998). *California's class-size reduction initiative: Differences in teacher experience and qualifications across schools (CS-006).* Riverside, CA: University of California, Riverside, California Educational Research Cooperative.

Reynolds, A. J., and Bezruczko, N. (1993). School adjustment of children at risk through fourth grade. *Merrill-Palmer Quarterly, 39*(4), 457-480.

Robelen, E. W. (1998). Reducing class size: When politics meet practice. *ASCD Infobrief*, Issue 14.

Veenman, S. (1995). Cognitive and noncognitive effects of multigrade and multi-age classes: A best-evidence synthesis. *Review of Educational Research, 65*(4), 319-381.

Wang, M. C., Haertel, G. D., and Walberg, H. J. (1993). Toward a knowledge base for school learning. *Review of Educational Research, 63*(3), 249-294.

Wexler, E., Izu, J., Carlos, L., Fuller, B., Hayward, G., and Kirst, M. (1998). *California's class size reduction: Implications for equity, practice & implementation.* San Francisco, CA: PACE-WestEd.

Wright, S. P., Horn, S. P., and Sanders, W. L. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education, 11*, 57-67.

Zykowski, J., Mitchell, D., Hough, D., & Gavin, S. (1991). *A review of year-round education research (YRE-002).* Riverside, CA: California Education Research Cooperative, School of Education, University of California, Riverside.